En *Moving and Interacting in Infancy and Early Childhood. An Embodied, Intersubjective and Multimodal approach to the Interpersonal World*. Cham (Suiza): Springer.

An invitation to embodied, multimodal and intersubjective approach.

Silvia Español, Martínez Mauricio y Rodriguez Fernando.

Cita:

Silvia Español, Martínez Mauricio y Rodriguez Fernando (2022). *An invitation to embodied, multimodal and intersubjective approach. En Moving and Interacting in Infancy and Early Childhood. An Embodied, Intersubjective and Multimodal approach to the Interpersonal World. Cham (Suiza): Springer.*

Dirección estable: https://www.aacademica.org/silvia.espanol/168

ARK: https://n2t.net/ark:/13683/pH0V/B1E

Acta Académica es un proyecto académico sin fines de lucro enmarcado en la iniciativa de acceso abierto*. Acta Académica fue creado para facilitar a investigadores de todo el mundo el compartir su producción académica. Para crear un perfil gratuitamente o acceder a otros trabajos visite: https://www.aacademica.org*.

Silvia Ana Español
Mauricio Martínez
Fernando G. Rodríguez *Editors*

Moving and Interacting in Infancy and Early Childhood

An Embodied, Intersubjective, and Multimodal Approach to the Interpersonal World

Springer

Moving and Interacting in Infancy and Early Childhood

Silvia Español
Mauricio Martínez
Fernando G. Rodríguez
Editors

Moving and Interacting in Infancy and Early Childhood

An Embodied, Intersubjective, and Multimodal Approach to the Interpersonal World

Springer

Editors
Silvia Español
Latin American Institute of Social Research (IICSAL)
Latin American Faculty of Social Sciences/ National Scientific and Technical Research Council (FLACSO/CONICET)
Buenos Aires, Argentina

Fernando G. Rodríguez
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Science (FLACSO), Buenos Aires, Argentina

Universidad Abierta Interamericana (UAI)
Buenos Aires, Argentina

Mauricio Martínez
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Science (FLACSO), Buenos Aires, Argentina

Universidad Abierta Interamericana (UAI)
Buenos Aires, Argentina

ISBN 978-3-031-08922-0 ISBN 978-3-031-08923-7 (eBook)
https://doi.org/10.1007/978-3-031-08923-7

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022
This work is subject to copyright. All rights are solely and exclusively licensed by the Publisher, whether the whole or part of the material is concerned, specifically the rights of reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed.
The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.
The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, expressed or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

This Springer imprint is published by the registered company Springer Nature Switzerland AG
The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Acknowledgment

We would like to thank Dr. Mario Carretero, General Director of the Department of Psychology of Knowledge and Learning at FLACSO Argentina, whose permanent support and encouragement made the publication of this book possible. We would also like to express our gratitude to the Department of Psychology of Knowledge and Learning at FLACSO Argentina for their financial support.

Contents

1 An Invitation to an Embodied, Multimodal, and Intersubjective Approach to Development 1
Silvia Español, Mauricio Martínez, and Fernando G. Rodríguez

2 The Infant-Directed Improvised Performances: What They Are and What Happens Through Them 21
Silvia Español, Favio Shifres, Isabel Martínez, and Diana Pérez

3 Infant-Directed Improvised Performances, Protoconversations, and Action Songs During the First Year of Life 57
Soledad Carretero, Silvia Español, Fernando G. Rodríguez, and Favio Shifres

4 Early Reciprocity: Temporal Coordination and Modality Behavior in Parental Imitation and Affect Attunement 91
Silvia Español, Mariana Bordoni, Mauricio Martínez, Santiago García Cernaz, and Viviana Riascos

5 Rhythm as a Prototypical Example of the Relationship Between Intersubjective and Perceptual Development During the First Year of Life 123
Mauricio Martínez and José Manuel Igoa

6 The Development of Multisensory Perception of Temporal Amodal Information 161
Mauricio Martínez

7 The Transition from Early Bimodal Gesture-Word Combinations to Grammatical Speech 207
Fernando G. Rodríguez and Silvia Español

8 Multimodal Expression in Communicative Functions, Gestures, Vocalizations, and the Contribution of Early Musicality 247
Fernando G. Rodríguez

9 Forms of Vitality Play 285
Silvia Español, Mauricio Martínez, Mariana Bordoni, Rosario Camarasa, Alicia Nudler, and Santiago García Cernaz

Index 329

Contributors

Mariana Bordoni Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Rosario Camarasa Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

Soledad Carretero Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Silvia Español Latin American Institute of Social Research (IICSAL), Latin American Faculty of Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina

Santiago García Cernaz National Scientific and Technical Research Council/ Laboratory for the Study of Musical Experience, Faculty of Fine Arts, National University of La Plata (CONICET/ LEEM-FBA-UNLP), La Plata, Argentina

José Manuel Igoa Department of Basic Psychology, Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain

Isabel Martínez Laboratory for the Study of Musical Experience, Faculty of Fine Arts, National University of La Plata (LEEM-FBA-UNLP), La Plata, Argentina

Mauricio Martínez Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

Alicia Nudler Laboratory of Research in Education of Teachers, Universidad de Río Negro, Bariloche, Argentina

Diana Pérez Philosophical Research Institute/SADAF-CONICET, Universidad de Buenos Aires, Buenos Aires, Argentina

Viviana Riascos Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

Fernando G. Rodríguez Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

Favio Shifres Laboratory for the Study of Musical Experience, Faculty of Fine Arts, Universidad Nacional de La Plata (LEEM-FBA-UNLP), La Plata, Argentina

Chapter 1
An Invitation to an Embodied, Multimodal, and Intersubjective Approach to Development

Silvia Español, Mauricio Martínez, and Fernando G. Rodríguez

Abstract This introductory chapter begins by explaining the origin of this book: a cooperative and interdisciplinary path walked by its authors that led to a conception of developmental psychology in permanent dialogue with the psychology of music, the philosophy of mind, and human movement studies. The overall framework of each chapter is then presented, arising from the confluence of three contemporary points of view: embodied cognition, the multimodal approach to child development, and the second-person perspective in social cognition. Such a framework allowed us to revisit some relevant topics in current developmental psychology in a new light, including adult–infant interactions, intersubjective experiences, the development of perceptual skills that allow infants to participate in the interpersonal world, verbal and gestural communication skills, and the complexity of play in infancy and early childhood. The chapter also provides an overview of each chapter in the book, outlining briefly their thematic connections.

Keywords Embodiment · Experience · Intersubjectivity · Second-person · Movement · Multimodal

S. Español (✉)
Latin American Institute of Social Research (IICSAL), Faculty of Latin American Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina
e-mail: sespanol@flacso.org.ar

M. Martínez · F. G. Rodríguez
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_1

1.1 About This Book

The authors of this book have a long history of working together on different research projects. Sustained cooperation over time, the exchange of ideas, and a permanent questioning of the assumptions that support theoretical formulations and empirical research in their respective fields of knowledge gave birth to a common conception of development and human life. In this conception, interpersonal encounters take center stage.

Certain topics of human experience are repeatedly explored, consistently receiving ineluctably partial answers. Developmental psychology has a rich history of research on such topics. It studies how the first interpersonal encounters emerge, how intimacy and intersubjectivity between adult and baby develop; it explores the development of perceptual abilities that allow babies to become involved in the interpersonal world; it deals with the unfolding of gestural and verbal communication skills; and it elucidates the complexity of play in infancy. Academic knowledge on these topics increased in the last decades, and new questions and research strategies emerged. Our Research Group on Socio-Cognitive Processes in Infancy and Early Childhood, at the Psychology of Knowledge and Learning Department in FLACSO (Facultad Latinoamericana de Ciencias Sociales/Latin American Faculty of Social Sciences) in Argentina, focuses on these topics. They are all addressed in one or more chapters of this book. Our approach to them is a product of interdisciplinary cooperation. This book was shaped through the meetings of our research group in developmental psychology, including what we consider the extended team – which includes members of SACCOM (Sociedad Argentina para las Ciencias Cogntivas de la Música/Argentine Society for the Cognitive Sciences of Music) and SADAF (Sociedad Argentina de Análisis Filosófico/Argentine Society of Philosophical Analysis). Based on these encounters, we generated a conception of developmental psychology in permanent dialogue with the psychology of music, the philosophy of mind, and human movement studies.

From these years of cooperation, emerged a shared conceptual framework that underlies each chapter. Within this framework, three viewpoints somewhat distant from the mainstream that currently guides most developmental psychology research converge: embodied cognition, the multimodal approach to infant and child development, and the second-person perspective on social cognition. These viewpoints are increasingly present in theoretical formulations and in observational and experimental empirical research designs. The first one, embodiment, arises as a response to the postulates assumed by classical cognitive psychology, which considered the body as a peripheral complement to apprehend the nature of the mind and cognition. Embodied psychology assumes that the body (not only the brain) plays a significant role in explaining the functioning and ontogenesis of psychological experience. The multimodal approach, on the other hand, reflects a particular way of understanding how people fit into their environment, which simultaneously provides their perceptual systems with diverse and often redundant information. Every perceptual system linking individuals to the world is functionally organized. As a result, they dynamize

the subject–world connection. Following the multimodal proposal, the history of our first relationship with the environment is rewritten, challenging a long-established tradition according to which the different perceptual modalities originally functioned in a disaggregated way (giving the newborn a chaos of unconnected stimuli). In our view, from the beginning of life, our environment is multimodal (as well as social). Therefore, the development of multisensory perception becomes an essential aspect for understanding the ontogenesis of specifically human psychological functions, both in their subjective and intersubjective aspects. The multimodal approach reinforces the contribution and novelty of embodiment because once again the body and its perceptual and motor patterns and possibilities are the key to understand human development. While not the same, embodied cognition and the multimodal approach go hand in hand to promote a different version of psychological development. Finally, the second-person perspective emerges as an alternative to classic cognitive psychology and philosophy where mental processes are considered opaque, and the understanding of one's own and other people's behavior requires the attribution of propositional content. Conversely, the second-person approach assumes that intersubjective contact depends on the direct perception of other people´s emotions and intentions and the development of reciprocity patterns. In this sense, intersubjective processes rely on early developmental perception-action processes, which remain the basic way of understanding others within an interactive context. In this initial format of interpersonal relationships, contacts are based on corporeal-affective exchanges between adults and infants. For this reason, the second-person perspective is also supportive of the embodiment approach. The three paradigms call upon and refer to each other, establishing a theoretical vantage point and promising horizons for empirical research.

1.2 An Embodied, Multimodal, and Intersubjective Approach to Development

The three viewpoints implied in our thinking about psychological development were investigated relatively independently and/or with subtle references to each other. Separately, each had a decisive renewing impact. Nonetheless, we are convinced that they function synergistically at the conceptual level and should therefore be considered together, both theoretically and empirically.

1.2.1 Embodied Experiences

The mind/body dichotomy of early cognitivism turned into mind/brain dichotomy, in line with advances in neuroscience. But reducing a whole corporeal being to the brain alone distorts human reality, the actual determinants of its behavior, and the

quality of subjective and intersubjective experience. In recent times, the mind–body linkage has been revisited anew from approaches emerging within cognitivism (understanding the mind as embodied, embedded, enacting, and extended – Newen et al., 2018; Rowlands, 2010). Each of these formulations gives a different value to the embodied experience.

The enactive program proposed by Varela et al. (1991) is probably the intellectual initiative that more openly addressed the experiential dimension of human life in cognitive science. This program's basic intuition is that our understanding is rooted in the structure of our biological embodiment but is lived and experienced through common actions and cultural history. Cognition depends on the experiences of being a body with many sensorimotor skills, embedded in a biological, psychological, and cultural context. It rises from repeated sensorimotor patterns that allow action to be perceptually guided. Motor and sensory processes cannot be separated from lived cognition. Cognition is understood as a world enactment – bringing about a world – through a viable history of structural coupling. From birth, we enter into a shared world, a grounding situation that allows us to make sense of our world or to *have a world*. Some of its current versions retain this original feature and emphasize that experience, far from being an epiphenomenon, is intertwined with being alive and immersed in a world of meanings. The shared world is constituted by the intersubjective understanding that shows things in the shared meaning they have for the members of a culture (Di Paolo et al., 2010; Durt et al., 2017). The enactive program also proposes a new approach to the neuroscience of consciousness and suggests that the processes crucial for consciousness cut across the brain–body–world division, rather than being brain-bound neural events. This approach allows for theories and hypotheses about the reciprocal relationship between embodied conscious states and local neuronal activity (Thompson & Varela, 2001).

For its part, embodied developmental psychology showed that the body is the main agent in the interactions shaping the mind of children. Development occurs through dynamic interactions between the brain, the body, and the physical–social–affective–cultural environment. As Thelen (2000a) pointed out, the mind is embodied: it emerges from and is continually influenced by bodily interactions. Consequently, developmental psychology not only acknowledges the sensorimotor origin of cognitive processes, but it is also committed to exploring the intimate and inextricable link between action and thought throughout life. What changes over the course of development is the nature of the coupling. What should therefore be investigated is not only the emergence of the structures of abstract cognition, but the emergence of a flexible and dynamic merge between direct online cognition and one less closely connected to the immediacy of the senses and the moving body. The more sophisticated cognition, organized in symbols or representations, will remain within the parameters it received from experiencing the world with a binocular vision, a stereophonic hearing, two free hands to reach objects, and an upright position for locomotion. Thinking starts with perceiving and acting, retaining its brand of origin forever. As Thelen (2000a, b, 2005) points out, the functional mapping of the brain relies on experience, especially the one coming from perception and motor examination. When babies are awake, they continuously move and search for things.

Every acquired experience adds to a network where paths are strengthened. The neural map is based on lived experience. Experience shapes the brain.

In this particular version of embodiment, the study of development emphasizes goal-directed movement, i.e., action. For example, Thelen et al. (1996) explored the genesis of the action of reaching. In a Piagetian vein, they were interested in movement-for-action related to the world and saw infants as real-time, context-specific problem solvers. Each baby in his/her own way, sooner or later, makes contact with a toy by tapping or rubbing it or sliding his/her hand over it. These contact moments involve certain movements, carving out increasingly recurring patterns. The cycle – excitement at the sight of a toy, action, and occasional contact – is repeated for weeks, until the efficient, successful action of reaching finally emerges. Reaching is the child's natural solution to obtaining a distal object and bringing it to the mouth. The path to successful action is not predictable. Infants differ in their initial conditions of body size and energy level, facing different postural and biomechanical problems. Each child finds a solution following individual developmental pathways that eventually converge in the successful action of grasping the object. Likewise, within the study of development within a dynamic systems framework, Smith (2013) argued in favor of the prospective importance of movement: movement may be involved in a complex, multilevel chain of events through which a new structure or function is formed. In this way, rolling to a sitting position allows children to hold objects and stabilize their heads. As they gain head movement control, they are able to place the focus of attention on the object while holding and manipulating it. This motor achievement has strong cognitive implications: it seems to be the best precondition for learning to name objects, even comparing with cases where the adult holds the target object, or cases where the object is close to the child, albeit not in his or her hands. Also, by manipulating the object, the child can see it from different spatial perspectives, allowing him/her to generate an abstract object shape, and from there, to substitute objects in pretend play. In this version of embodiment, motor development is understood as the ideal model system for the study of psychological development (Thelen, 2000b; Adolph & Hoch, 2019).

But the embodied experience is broader and deeper. Regardless of their assertiveness, these theories consider movement either at the service of action development or at the service of cognitive development, both depending on what movement allows to happen in the future. Moreover, they hardly consider the affective world, and if so, it is usually reduced to basic emotions. Darwin (1872) introduced evolutionary considerations in order to understand expressions of the emotional world. He assumed that the habit of expressing feelings through certain movements would have been naturally selected in each species and defended the existence of a reduced set of universal emotional expressions in human species. Ekman and Friesen (1971) and Izard (1979) resumed the Darwinian idea and brought the study of the expression and interpretation of emotions to the laboratory, confirming the existence of a few basic and universal emotions – anger, fear, joy, disgust, sadness, interest, surprise – each with their own discrete and innate facial display. From then on, the study of basic emotions was always present in cognitive psychology research, but also, to our knowledge, it generally focused on a relatively static view of emotions

(Despret, 2004), with few exceptions (Damasio, 2003). It never reached the dynamic aspect of the affective experience.

The tight limits imposed by a discrete and static view of the emotion were overcome by the recognition of the central role of time in subjective and intersubjective experience. The temporal dimension, in this case on the ontogenetic and microgenetic levels, made its way into the first studies of early interactions and the description of adult–infant communion states (see the volume edited by Schaffer, 1977). The methodological changes were immediate: microanalyses were added to the longitudinal studies in this field, endowing the observation of early interactions with a temporal magnifying glass. In this way, the organization of adult–infant encounters in seconds and milliseconds gradually became visible. Along with this, of course, the narrow view that early cognitive psychology held of the human affective world, which tended to reduce all affective life to the Darwinian repertoire of basic emotions, was expanded.

> When thinking about the experience of young infants, one frequently falls back on the metaphors of music. This is hardly strange, given that both types of experiences are nonverbal, intuitively grasped, and involve the flow of forms in time. I will suggest that certain basic experiences of time and form that are common to our encounter with music are also common to an infant's ordinary, daily, socio-affective interactions (Stern, 2000, p. 21).

Decades later, studies in music psychology began to unveil infant sensitivity to musical qualities (Papoušek, 1996). Communicative musicality research (Malloch & Trevarthen, 2009) also brought musical timing (in terms of pulse, quality, and narrative) into the domain of affects and feelings traversed by the adult–infant dyad. Like in several previous studies in early interactions, the multimodal character of adult–infant encounters was highlighted. And musical variables (such as timing, melodic contours, or timbral variations) were included in the microanalyses of adult–infant interactions.

This stream of thought opened the doors to an aesthetic view of development that paid attention to a temporal interaction dimension whose flow and meaning are similar to the musical phenomenon. Subsequently, Dissanayake (2000, 2008, 2017), emphasizing a multimodal perspective of adult–infant interaction, stressed that antecedents of temporal arts, not only music but also dance, can be found in them. As Sheets-Johnstone (2011, 2015) indicated from a philosophical approach, the mind/body dichotomy could only be overcome by granting primacy to movement because thinking and moving are not separate events but aspects of a kinetic bodily logos attuned to an evolving dynamic situation.

Stern, who always maintained a multimodal view on development, introduced an aesthetic understanding of movement through the concept of vitality forms. Stern (2010) defines the experience of movement as the original source of psychological life. Movement provides a primary sense of vitality, from which *vitality forms* emanate. The experience of vitality is the result of a dynamic pentad composed of movement and time, force, space, and the direction/intention it carries. Forms of vitality are directly perceived in our kinesthetic experiences and can be directly observed in the behavior of others; they concern the how, the style of doing things. Perceiving

the forms of vitality in oneself and in others is an ability we acquire at the beginning of life that marks our psychological life throughout our life cycle. Stern's description of the forms of vitality recalls Laban's (1971) – choreographer and theorist of modern dance – concept of effort (energy/quality). Effort is determined by a surrendering or struggling attitude toward four factors: weight, space, time, and flow. Thus, a movement can be firm or light, focused or dispersed, sudden or sustained, and free or driven in its multiple degrees and combinations. Stern (2010) carries the vision of movement in dance to the center of psychological life: because the forms of vitality permeate all our movements, actions, and behaviors and because the heart of the interpersonal world lies in the expression and awareness of the forms of vitality. By paying attention to the interoceptive, proprioceptive, and kinesthetic qualities of infant experience and their states of enhanced connectivity or resonance with others, it is possible to describe how the mapping of infant embodied awareness develops. Children's movements reveal the corporeality of their culture. From very early on, they incorporate forms of movement distilled from their culture, such as the way they move and dance. These enculturated movements can sometimes be used for communicative purposes and can also become a symbolic action (Español, 2007). As Negayama et al. (2015) point out, the timing, form, and energetic of body movements can be specific to a culture and learned in early adult–infant engagement. The feelings conveyed in body movement, in the choice or form of action, form a basis of cultural knowledge and evolution. Foregrounding the present perceptual/kinetic bodily experience does not deny our symbolic nature but constrains its dominance (Español, 2018). This formulation of embodiment, which gives rise to an aesthetic vision of development and foregrounds the experience of movement, is what characterizes our work.

1.2.2 Intersubjectivity and Second-Person Perspective

The very early emergence of intersubjective experiences in the infant–adult dyad is connected with the phylogenesis of our species. As Bruner (1972) pointed out, primate evolution relies on the progressive selection of a characteristic immaturity pattern whose main effect was the emergence of a more flexible form of social bonding. Human primates ride the crest of the wave of this evolutionary selection by having the longest immaturity period (Gómez, 2004). Our prolonged period of immaturity is probably the most impressive biological fact of our ontogeny (Bruner 1996). The dependence and the omnipresent paternal scaffolding from birth derive from immaturity and essentially affect the human social world (Gómez, 2004; Rochat, 2001). The conjunction of infant helplessness and parental scaffolding seems to have paved the way for one being (helpless) and another being (caregiver) to engage in intimate psychological contact experiences that are a necessary condition for human development.

Trevarthen (1982, 1998) called these experiences of intimate psychological contact intersubjectivity and distinguished between primary and secondary subjectivity.

Primary intersubjectivity is the most primitive mode of psychological connection where each subjectivity orients itself to the other; it is the moment of dyadic interactions, involving shared emotional states, where objects of the external world are still absent in infant–adult exchanges. Secondary intersubjectivity, on the other hand, implies the coordination of self, other, and object. Its first manifestations are the phenomenon of joint attention, where infant and adult alternate their gaze between an object of interest and her/his partner, and the first communicative gestures used by the infant to direct adult attention. Primary intersubjectivity experiences emerge around the age of 2 months, and the transition to secondary intersubjectivity consolidates once the capacity for joint attention and action toward objects in the environment becomes evident, around the age of 9 months. How and when this transition from one to another occurs remains a subject of controversy and exploration. Bråten and Trevarthen (2007) recognized a tertiary level of intersubjectivity involving symbolic references, the use of metarepresentations, and narrative speech. The latter emerges between the ages of 2 and 4 years, approximately when children gradually begin to engage in conversations and represent interactions with others in terms of intentions and reasons, recognizing different perspectives.

While there is certainly an ontogenetic sequence from primary and secondary intersubjectivity to the later mode of tertiary intersubjectivity, it does not necessarily imply the disappearance of the earlier forms. Assuming a nonlinear conception of development, the earlier stages do not disappear but remain reconfigured. For example, Stern (1985) described four senses of the self, emerging between birth and the age of two, which depend on the infant's relationship with himself, with others, and with objects. Each sense of self (emergent, nuclear, subjective, and verbal) brings about new organizing perspectives of the self and others that instead of disappearing are integrated into the following and can still be fully experienced. We are interested in the experiences of primary and secondary intersubjectivity and the ways they perdure throughout development. We are also attentive to processes that, while preserving essential features of the first intersubjectivity modalities, orient themselves toward tertiary intersubjectivity, especially in communication and play.

Currently, the proposal that comes closest to the nonlinear way of understanding the development of intersubjective processes is the second-person perspective in social cognition. This perspective emerges as an alternative to the answers given by classical cognitivism to the problem of other minds – a problem that refers to the human experience of being irretrievably linked and at the same time opaque to each other. The traditional first- and third-person approaches in social cognition emphasize the experience of distance and opacity of the internal states of others: the third-person perspective adopts a rationalistic and deductive view of the existence and nature of other minds, and others are seen as a distant she or he; the first-person perspective traces our awareness of others to the self's direct experience and to its projected simulation on others (something is experienced in the first-person's private domain and then extended to others). From a renewed first-person perspective focused on embodied simulation, mirror neurons and mirror-related mechanisms were proposed as the neurobiological grounding of some forms of primary and secondary intersubjectivity (Ferrari & Gallese, 2007). Mirror neurons and other

embodied mechanisms are considered to support the aesthetic experience in dance (Calvo-Merino, 2010) and in the visual arts (Freedberg & Gallese, 2007; Ferrari & Gallese, 2007). These findings endorse the idea that we are oriented or driven to feel sympathetically certain movement qualities of other people. But this does not necessarily imply involvement with others. We may feel sympathetically certain qualities of the movement of other people whom we observe from a distance while not engaging in actual bodily contact.

As Reddy (2008a) points out, in these two traditional perspectives, the question of how humans become aware of the psychological qualities of people proceeds from the postulate of the existence of a gap, a deep hiatus, between the self and the other. But the experience of our relationships with others does not end with the perception of distance. We equally or more clearly experience closeness and intimacy, the transparency of our inner worlds. By emphasizing closeness, we can observe that the intimate and transparent encounters of our worlds are bodily encoded. The second-person perspective replaces the idea of a distant and inescapable opacity between us with the image of a perceptual union of active reciprocity (Reddy, 2008a). While traditional perspectives assume we have a first-person perceptual experience of our mental life that stands apart from the distal third-person perception of the behavior of others, the second-person perspective stresses that the perception of others consistently involves proprioception. Being in contact with another person presupposes an assembling between exteroception and proprioception. Perceiving someone directing their attention to us inescapably involves the proprioception of our own response. For infants, the experience of being noticed by others is one of the earliest forms of having a direct experience of the other (Reddy, 2005 and 2008b). Being looked at is, perhaps, the most dramatic case of proprioceptive information provided by the perception of others, the strength of which does not diminish throughout life. From the second-person perspective, the center of our interpersonal world is the reciprocal, rapid, and dynamic exchange. This is the essential difference between a renewed first-person perspective focused on embodied simulation and the second-person perspective. The ways of *being with* the other are based on a set of skills for reciprocal understanding that develops and expresses within interactive contexts. What a subject does elicits a sensitive response in the other; actions and reactions involve reciprocity. The expressive aspects of the behaviors of others – such as voice tone, smile, or muscle tone – are perceived to be directly significant and constitute the basis for a corresponding reaction fully meaningful within the interaction (Gomila, 2002, 2003; Scotto, 2002; Thompson, 2001; Gomila & Pérez, 2017). The second-person perspective is constitutively embodied (Pérez & Gomila, 2022). Moore and Barresi (2017) distinguish five forms of second-person information – self-directedness, contingency, reciprocity, affective engagement, and shared intentional relations – that provide the critical substrate for the development of social understanding.

The second-person perspective exhibits the direct, transparent, I–You contact established in the first adult–infant reciprocity patterns. We can perceive each other's feelings and intentions in their postures, movements, facial expressions, gestures, vocal utterances, and actions and engage in cycles of reciprocity. Although

these experiences appear early in development, they continue to be the basic way of understanding others in interactive situations (Gallagher, 2001, 2008; Gallagher & Zahavi, 2008).

According to Pérez and Gomila (2022), second-person interactions typically involve two people interacting directly within a given context. For example, the interactions between an infant and a caregiver, or tango dancers in a milonga (where there is no predesigned choreography but only two people mastering a technique and joining in a flowing melody) are both cases of genuine second-person interactions. Tango dancers clearly abide by cultural norms; what is not always highlighted is that cultural norms are also involved in the interactions between a caregiver and an infant.

> Second person interactions are basic but they always incorporate many - explicit or implicit cultural norms and knowledge, which can differ among different social groups (….) Cultural norms and practices … are always integrated into human interactions, even in the case of a newborn baby and her/his caregiver: no human interaction is alien to normative and cultural settings (Pérez & Gomila, 2022, p. 4).

The second-person perspective places the body-present encounter and reciprocity patterns at the origin and center of human ways of being with one another. But "(…) most of our second person encounters in real life involve the interplay of the three perspectives" (Pérez & Gomila, 2022, p. 109).

The adoption of the second-person perspective in developmental psychology induces us to rethink our research methods. Reddy (2008a) proposes a second-person methodology for psychological research on intersubjective phenomena, as opposed to distanced, objective third-person studies and subjective first-person studies. She suggests that the knowledge we can generate concerning interpersonal involvement phenomena essentially depends on the intersubjective relationship that is established, i.e., on the genuine person-to-person encounter. Thus, the second-person perspective contemplates the option of a researcher who is included as an active participant.

1.2.3 The Multimodal Interpersonal World in Infancy and Early Childhood

Infant psychological development occurs in an eminently multimodal social context. Caregivers engage their infants in multiple activities: they feed them, relieve their bodily discomforts, bathe them, regulate their attentional and emotional states, and devote time to play together freely. The systematic observation of these activities reveals a common aspect: multimodality. In every child-rearing action, adults are inclined to provide simultaneous information to the infant's different perceptual systems.

The concept of forms of vitality (Stern, 2010) – covered in most of the chapters in this book – permits a unified qualification of bodily, mental, and inter-subjective

experiences in their temporal becoming. As E. Gibson (1969) once pointed out, one of the most interesting characteristics of temporal experience is that some of its defining features can be perceived through different perceptual systems. These characteristics are often referred to as amodal information. Duration, rate, and rhythm are some of the qualities that are perceived interchangeably across different perceptual systems. The temporal unfolding of the infant's intersubjective encounter with her or his parenting figures may be characterized according to the multimodal modeling of the duration, rate, or rhythm of the information provided by the adult. Contemplating this aspect of infant experience reveals the need to understand the multisensory perceptual abilities of the infant and their development during the first year of life.

According to Bremner et al. (2012), most contemporary theoretical formulations tend to approach development from largely unimodal theoretical frameworks. Likewise, the scant attention paid to some perceptual systems of vital importance to infant development, including touch (Field, 2014), is also highlighted. However, in recent years, the recognition of the relevance conferred to multisensory perception has increased considerably. For example, Bahrick et al. (2020) argue that multisensory perception serves as a critical foundation upon which more complex social, cognitive, and language skills can develop.

Explanations of the development of multisensory perception – e.g., Bahrick and Lickliter (2012), Bahrick et al. (2020), Gibson and Pick (2000) – do not see intersubjective experience as a shaping element. However, as proposed by Rivière long ago, the relationship between multisensory perception skills and interpersonal relationship skills "should be posed in dialectical terms rather than in terms of a linear determination […] interaction itself seems to contribute to the development of the ability to assimilate information from different sensorial origins" (Rivière, 1986/2003, p. 114). Therefore, the idea to be explored is how infant–caregiver interaction facilitates the development of multisensory perception while the development of multisensory perception impacts their interaction.

The intersubjective encounter of infants with their caregivers is extensively multimodal. It involves exteroceptive, proprioceptive, and interoceptive experiences. The parents offer the infants simultaneous information for their different perceptual systems. They touch, talk to, caress, and move the infant. All this information is usually temporally patterned: the duration of the caresses the adult makes tends to match the duration of her/his vocalization, and the rhythmic pattern by which the adult taps her/his palm on the abdomen of the infant matches the rhythmic pattern of vocalizations. In this way, the infant receives redundant sensory information through two, or more, perceptual systems. Recent studies have exposed the immense value of this type of information for infant development (Bahrick et al., 2020).

The information provided by the adult in the above examples may create the impression that intersubjective encounters ultimately consist of a sound and movement spectacle where infants are mere observers. Their experience, however, transcends the distant experience of the adult as a perceptual frame. Shifts in the temporal dynamics of adult information coincide with changes in the bodily sensations of the baby, in his breathing, in his muscular tension or relaxation, and so on.

As we indicated in the previous section, intersubjective encounters presuppose an assembling between exteroception and proprioception. An interesting experience of such assembling is affect attunement (Stern, 1985). Adults often affectively attune to infants. For example, as the infant rhythmically and repeatedly hits a toy on the floor in front of his mother, she accompanies each movement with vocalizations matching the rhythmic pattern of the infant's strokes. The coincidence between exteroceptive rhythmic information (the hearing of maternal vocalizations and perhaps of the toy hitting the floor) and proprioceptive and interoceptive information (the sensation of his body moving and the vibration of the hitting on his arm) provides the infant with a unique intersubjective experience where he shares the same form of vitality he experiences in herself/himself with a partner.

Each chapter in this volume addresses the multimodality characterizing the world of infants and young children. Most of them also reflect the link between intersubjective and perceptual development. In this way, they each contribute to the generation of increasingly close links between the three approaches that make up the framework that guides our understanding and inquiry into development.

1.2.4 Our Theoretical and Empirical Approach

As we stated at the beginning, embodied cognition, the multimodal approach to development, and the second-person perspective on social cognition, although conceptually linked, can and do develop independently. In exploring some interactive psychological processes relevant to human development, we find that the confluence of these three viewpoints allows us to revisit, in a renewed light, relevant topics of development. An original feature of this book consists in providing a framework where the three points of view combine while also linking them by using concepts and analytical categories from the temporal arts – music and dance. This way of conducting empirical research on some central topics in early infancy led us to an aesthetic conception of development that emphasizes bodily experience and temporal affects while investigating their intertwining with our symbolic capacities.

A final particularity of this book is that, within the aforementioned framework, a flexible and plural methodological stance is adopted, allowing a convergence between studies with different designs. We performed longitudinal case studies where, following second-person methodology guidelines, the researcher is included as a member of the studied family or as an active participant, establishing a warm relationship with the caregiver and the child. Although this type of study may find limitations when generalizing results, it allows a detailed follow-up of the dyad and closely reflects the ontogenetic developmental process. We frequently select interaction scenes from our longitudinal studies to carry out sound and movement microanalyses using concepts from dance and music, as well as analytical categories from music performance studies and the Laban system of movement analysis. These studies afford a "temporal magnifying glass" to observe the microgenesis of interaction. Microanalysis provides information that naturalistic observation in real-time

cannot give. In Valsiner's (2014) words, research methods are selected according to the researcher's conception of development and the desired level of genesis to be observed. When research questions require it and the state of the art allows it, we resort to conducting experiments. In these experiments, we take the greatest care to ensure that the stimuli used are ecologically valid. Theoretical inquiry is always a constant horizon in our work, connected with empirical research. In our almost artisan way of doing psychology, detailed empirical observation functions as a trampoline to theoretical debate.

1.3 Overview of the Chapters

When adults interact with babies, they do special things. In Chap. 2, *The Infant-Directed Improvised Performances: What They Are and What Happens Through Them*, Silvia Español, Favio Shifres, Isabel Martínez, and Diana Pérez describe what they have called *infant-directed improvised performance* through a microanalysis of interaction scenes between an adult and a 7-month-old baby. The infant-directed improvised performances are sound-kinetic phrases improvised by the adult with resources also used in temporal art performances, like the repetition-variation form. Adults create brief motifs collected from the behavior of the baby or from contingencies in the surroundings and repeat them, varying their form. The authors employ analytical tools developed for the exegesis of artistic expression. They use the Laban system of movement analysis and analytical tools in musical interpretation to describe the temporal, energetic, and spatial dimensions of motif variations. Infant-directed improvised performances lengthen the intersubjective moment, allowing infants to experience a range of "forms of vitality" (a temporary affect not reducible to classical Darwinian emotions) that they could not access on their own. Through infant-directed improvised performances, adults summon them into social life, offer them well-formed units of behavior that favor their recognition, and interpret the world for infants, regulating their moods. Adults also illuminate cognitive structures such as image schemas and invite infants to experience primary metaphors. By virtue of performances, babies are also initiated into corporeal and aesthetic enculturation.

The connection between babies and adults is crossed by what has been called "communicative musicality" (a mode of intersubjective encounter with features common to the temporal arts). In exchanges between adults and infants, this encounter mode is manifested in different ways. In Chap. 3, *Infant-Directed Improvised Performances, Proto-Conversations, and Action Songs During the First Year of Life*, Soledad Carretero, Silvia Español, Fernando G. Rodríguez, and Favio Shifres analyze three variants of corporeal and intersubjective encounters relevant to social cognitive development: the aforementioned infant-directed improvised performances, protoconversations, and action songs. The three form part of the first social and multimodal exchanges that allow adults and babies to share moments of deep intimacy and affection. While protoconversations are extensively studied

interactions in developmental psychology and action songs are widely known, infant-directed improvised performances have only recently been identified and were outlined in the previous chapter. The three variants of these exchanges are brought together as a unitary category under the name of *emergent organizations of early communicative musicality*. The chapter presents a longitudinal case study of a mother–infant dyad between the 2nd and 9th month of the baby, analyzing the structure, frequency, and developmental trajectory of these three emergent organizations. It also describes the various ways in which multimodal stimulation is organized in each of them. Chapters 2 and 3 can be considered together as an exploration of infant-directed impromptu performances.

The interpersonal world assumes a rapid, dynamic, reciprocal, body-present exchange with whom we interact. Adults are responsive to infant behavior from the first months of life and thus scaffold their learning of interpersonal reciprocity. Certain adult responses involve some kind of infant behavior matching. Maternal imitation and affect attunement are two recognized matching activities that set a special degree of intersubjectivity and mutuality. In Chap. 4, *Early Reciprocity: Temporal Coordination and Modality Behavior in Parental Imitation and Affect Attunement*, Silvia Español, Mariana Bordoni, Mauricio Martínez, Santiago García Cernaz, and Viviana Riascos present a longitudinal case study of a mother–infant dyad where these two matching activities are analyzed between the infant's 2nd and 10th month of life. The focus was set on the development of maternal imitation and affect attunement frequencies, the temporal coordination between infant behavior and adult response, and the pairing between infant and maternal behavioral modalities (vocalization, facial expression, movement). Some results proved noteworthy: attunements were more frequent than imitations, and affect attunement patterns were mainly based on maternal receptivity and reciprocity toward infant movements. Maternal responsiveness matching activities were also found to be temporally coordinated according to very short time periods (a time window of less than one second). Two predominating coordination modes were described: overlapped affect attunements in response to infant movement; and maternal imitation with a switching pause in response to infant vocalizations.

In Chap. 5, *Rhythm as a Prototypical Example of the Relationship Between Intersubjective and Perceptual Development During the First Year of Life,* Mauricio Martinez and José Manuel Igoa propose rhythm as a paradigmatic example of the relationship between intersubjective development and the development of multisensory perception during the first year of life. Embodied cognition and the second-person perspective allow us to think of perception as a psychological process that enables social exchange. The specialized literature does not usually specify which perceptual abilities might be involved in the perception of other people's behavior. Under the assumption that in infant intersubjective development there are changes in the perceptual capacities involved in the transition from primary to secondary intersubjectivity, this chapter links the development of multisensory rhythm perception to infant intersubjective development and describes the ontogenesis of multimodal rhythmic pattern detection, discrimination, and recognition. The authors review the empirical evidence on these perceptual skills and give a detailed

description of an experimental study conducted by Mauricio Martínez, Silvia Español, and José Manuel Igoa on rhythm-based intersensory relationship perception development. The peculiarity of this study lies in the type of stimulus used: a video of a woman producing rhythmically modeled sounds and movements that mimic an improvised performance led by a child. Considering the findings along with evidence from other studies, the authors empirically and conceptually outline the development of the three multisensory rhythm perception skills linked to intersubjective development during the first year of life.

In Chap. 6, *The Development of Multisensory Perception of Temporal Amodal Information,* Mauricio Martínez presents a theoretical model of the ontogenesis of multisensory perception of temporal amodal information (MPTAI). The development of MPTAI involves the three different perceptual skills detailed above (detection, discrimination, and recognition) linked to amodal temporal information (duration, speed, and rhythm). The model assumes a dialectical relationship between early intersubjective development and the development of amodal temporal information. The model provides an ecological approach to perception integrating a social–historical–cultural framework, relying on the genesis of infant exploratory activity. This exploration becomes the source of the amodal temporal information in adult behavior that the infant learns to extract to shape his or her relationships with others. Chapters 5 and 6 can be seen as two sides of the same coin. While Chap. 5 (taking rhythm as an example) focuses on establishing relationships between the development of three multisensory perception skills of amodal information (detection, discrimination, and recognition) and the intersubjective development of the infant, Chap. 6 emphasizes the explanation of the development of these skills in terms of infant participation in intersubjective encounters with their nurturing figures.

Concerning the first signs deliberately addressed to adults, children's repertoire contains gestures, single words, and also bimodal combinations of orality and gestuality**.** Chapter 7, *The Transition from Early Bimodal Gesture–Word Combinations to Grammatical Speech*, authored by Fernando G. Rodríguez and Silvia Español, presents some considerations on the status of gesture in infancy and early childhood, individualizing examples usually disregarded and on bimodal combinations, for which an observation code is proposed. This code is the product of a longitudinal case study inspired by the second-person methodological perspective. It resumes previous research approaches but pays special attention to bridging the division between gesture development and children's grammatical speech studies. It is organized around a relational-semantic axis to bring continuity between the study of pre-grammatical and grammatical communicative utterances. Tracking the use of signs in babies and young children from their single forms, through early combinations or juxtapositions until the first morphosyntactic speech, this approach would support a comprehensive understanding of the meaningful-functional development of communication devoid of abrupt discontinuities. The chapter also deals with the pending challenge of expanding the margins of bimodal combinations to include each multimodality form and variant of posture, movement, orality, etc., which, when contextualized, make up episodic meaningful units. Following the spirit of the

triple viewpoints in this book, which integrates corporeality, multimodality, and intersubjectivity as an exploratory-theoretical device suited to the addressed phenomena, the authors emphasize the need to move in the direction of a sensitive ecological code that includes all types of expressions, even normally meaningless behaviors.

Chapter 8, *Multimodal Expression in Communicative Functions: Gestures, Vocalizations, and the Contribution of Early Musicality*, authored by Fernando G. Rodríguez, addresses communicative functions in child expression. Pre-grammatical speech babies and young children use signs to interact with others according to their particular interests. These functional motives have certainly been the subject of diverse classifications, generally focusing on one modality (gesture, word), bimodal gesture–word combinations, or oriented by a biased criterion (e.g., in terms of types of signs, or illocutionary utterances, etc.). Aiming to be inclusive, this chapter proposes to classify communicative functions of oral, gestural, and combined modalities, also suitable for other types of signs and specially designed to deal with pre-grammatical expressive resources. The category system was elaborated considering both the structural scheme of communicative acts and the interactional intentions that lead pre-grammatical children to signify for others. Based on this double consideration and on the records of a longitudinal adult–baby dyad case study between the ages of 19 and 29 months of the infant, the observation code includes categories regularly absent in competent speaker classifications (adults, older children), excludes others requiring the emergence of certain earlier cognitive skills, and individualizes three new ones. Two of them are liminal to the communicative phenomenon and are linked to musicality and multimodality, allowing us to conceive children's semiotic development as a continuum from pre-communicative dyadic behavioral skills to communicative ones. Chapters 7 and 8 can be considered as a whole, a combined essay to contribute to the research on the semantic and pragmatic aspects of children's expression.

Chapter 9, *Forms of Vitality Play*, authored by Silvia Español, Mauricio Martínez, Mariana Bordoni, Rosario Camarasa, Alicia Nudler, and Santiago García Cernaz, is directly associated with Chap. 1. It presents a non-figurative play involving the child's appropriation of the improvisational resources used by adults in infant-directed improvised performances. Since the pioneering studies in psychology, play has been considered a distinctive feature of childhood. To this day, play continues to be the subject of a steady stream of research, but it remains predominantly focused on social and symbolic play. The authors review a longitudinal case study of an adult–infant interaction during the third year of life (guided by the second-person methodological perspective) where one type of play caught their attention. It has been dubbed *forms of vitality play*, in recognition of Daniel Stern's concept of *forms of vitality*. In *infant-directed improvised performances*, parents model forms of vitality through the repetition and variation of sounds and movements, while the baby is an involved recipient who participates with general social behaviors. During the child's third year of life, adult and child play together with forms of vitality, creating and embellishing sound and movement motifs according to the repetition-variation form, but now the child actively engages in the elaboration of the spatial,

energetic, and temporal dimensions of the motifs. Forms of vitality play is sometimes combined with symbolic play. The aesthetic perspective adopted in this book was the ideal tool for the authors to bring out, in a new light, the ability of children to playfully improvise in interpersonal contexts; such perspective allowed them also to demonstrate how children sometimes make use of play with the forms of vitality as a resource for the elaboration of symbolic events in pretend play.

References

Adolph, K. E., & Hoch, J. E. (2019). Motor development: Embodied, embedded, enculturated, and enabling. *Annual Review of Psychology, 70*, 141–164. https://doi.org/10.1146/annurev-psych-010418-102836

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–205). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Bahrick, L. E., Lickliter, R., & Todd, J. T. (2020). The development of multisensory attention skills: Individual differences, developmental outcomes, and applications. In J. J. Lockman & C. S. Tamis-LeMonda (Eds.), *The Cambridge handbook of infant development* (pp. 303–338). Cambridge University Press.

Bråten, S., & Trevarthen, C. (2007). Prologue: From infant intersubjectivity and participant movements to simulation and conversation in cultural common sense. In S. Bråten (Ed.), *On being moved: From mirror neurons to empathy* (pp. 21–34). John Benjamins Publishing Company.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). The multisensory approach to development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 1–26). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0001

Bruner, J. S. (1972). Nature and uses of immaturity. *American Psychologist, 27*(8), 687–708. https://doi.org/10.1037/h0033144

Bruner, J. (1996). *The culture of education*. Harvard University Press.

Calvo-Merino, B. (2010). Neural mechanisms for seeing dance. In I. B. Bläsing, M. Puttke, & T. Schack (Eds.), *The neurocognition of dance: Mind, movement and motor skills* (pp. 153–176). Psychology Press.

Damasio, A. (2003). *Looking for Spinozza: Joy, Sorrow, and the Feeling Brain*. [En busca de Spinoza: Neurobiología de la emoción y los sentimientos (J. Ros, Trad). Madrid: Crítica, 2005]. Harcourt.

Darwin, Ch. (1873/1965). *The expression of emotions in animal and man* [La expresión de las emociones en los animales (T. Fernández, Trad.). Madrid: Alianza Editorial, 1998]. The University of Chicago Press.

Despret, V. (2004). *Our emotional makeup*. Other Press, LLC.

Di Paolo, E., Rohde, M., & De Jaegher, H. (2010). Horizons for the enactive mind: Values, social interaction, and play. In J. Stewart, O. Gapenne, & E. Di Paolo (Eds.), *Enaction: Toward a new paradigm for cognitive science*. MIT Press.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389–410). The MIT Press.

Dissanayake, E. (2001). Becoming Homo Aestheticus: Sources of aesthetic imagination in mother-Infant interactions. *SubStance, 30*(1/2), 85–103. https://doi.org/10.2307/3685506

Dissanayake, E. (2008). The arts after Darwin: Does art have an origin and adaptive function? In K. Zijlmans & W. van Damme (Eds.), *World art studies: Exploring concepts and approaches* (pp. 241–263). Valiz.

Dissanayake, E. (2017). Roots and route of the artification hypothesis. *AVANT. Pismo Awangardy Filozoficzno-Naukowej, 1*, 15–32.
Durt, C., Fuchs, T., & Tewes, C. (Eds.). (2017). *Embodiment, enaction, and culture: Investigating the constitution of the shared world*. MIT Press.
Ekman, P., & Friesen, W. (1971). Constants across cultures in the face and emotion. *Journal of Personality and Social Psychology, 17*(2), 124–129. https://doi.org/10.1037/h0030377
Español, S. (2007). Time and movement in symbol formation. In J. Valsiner & A. Rosa Rivero (Eds.), *The Cambridge handbook of socio-cultural psychology* (pp. 238–256). Cambridge University Press.
Español, S. (2018). Early infancy: A moving world. Embodied experience and the emergence of thinking. In J. Valsiner & A. Rosa Rivero (Eds.), *The Cambridge handbook of socio-cultural psychology* (2nd ed., pp. 207–222). Cambridge University Press.
Ferrari, P., & Gallese, V. (2007). Mirror neurons and intersubjectivity. In S. Bråten (Ed.), *On being moved: From mirror neurons to empathy* (pp. 73–88). John Benjamins Publishing Company.
Field, T. (2014). *Touch* (2nd ed.). MIT Press.
Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Science, 11*(5), 97–203. https://doi.org/10.1016/j.tics.2007.02.003
Gallagher, S. (2001). The practice of mind. Theory, simulation or primary interaction? In E. Thompson (Ed.), *Between ourselves: Second person Issues in the study of consciousness* (pp. 83–108). Imprint Academica.
Gallagher, S. (2008). Understanding others: Embodied social cognition. In P. Calvo & A. Gomila (Eds.), *Handbook of cognitive science: An embodied approach* (pp. 439–452). Elsevier.
Gallagher, S., & Zahavi, D. (2008). *The phenomenological mind: An introduction to philosophy of mind and cognitive science*. Routledge.
Gibson, E. (1969). *Principles of perceptual learning and development*. Appleton-Century-Crofts.
Gibson, E., & Pick, A. (2000). *An ecological approach to perceptual learning and development*. Oxford University Press.
Gómez, J. C. (2004). Apes, monkeys, children and the growth of mind [El desarrollo de la mente en los siminos, los monos y los niños (B. Martín-Andrade, Trad.). Madrid, Morata, 2007]. Harvard University Press.
Gomila, A. (2002). La perspectiva de segunda persona de la atribución mental [The second-person perspective on mental attribution]. Azafea. *Revista de Filosofía, 4*, 123–138. https://doi.org/10.14201/3719
Gomila, A. (2003). La perspectiva de segunda persona de la atribución mental [The second-person perspective of mental attribution]. In E. Rabossi & A. Duarte (Eds.), *Psicología cognitiva y filosofía de la mente [Cognitive psychology and philosophy of mind]* (pp. 195–218). Alianza Editorial.
Gomila, A., & Pérez, D. (2017). Lo que la segunda persona no es [What the second person is not]. In D. Pérez & D. Lawler (Eds.), *Las emociones y la segunda persona. [Emotions and the second person]* (pp. 275–297). SADAF.
Izard, C. (1979). *The maximally discriminative facial movement coding system*. University of Delaware Instructional Resources Center.
Laban, R. (1971). *The mastery of movement*. Plays.
Malloch, S. & Trevarthen, C. (2009) Musicality: Communicating the vitality and interests of life. En S. Malloch and C. Trevarthen, (Eds). *Communicative musicality: Exploring exploring the basis of human companionship,* (pp. 1-10) Oxford University.
Moore, C., & Barresi, J. (2017). The role of second-person information in the development of social understanding. *Frontiers in Psychology, 8*, 1667. https://doi.org/10.3389/fpsyg.2017.01667
Negayama, K., Delafield-Butt, J. T., Momose, K., Ishijima, K., Kawahara, N., Lux, E. J., Murphy, A., & Kaliarntas, K. (2015). Embodied intersubjective engagement in mother–infant tactile communication: A cross-cultural study of Japanese and Scottish mother–infant behaviors during infant pick-up. *Frontiers in Psychology, 6*, 66. https://doi.org/10.3389/fpsyg.2015.00066
Newen, A., De Bruin, L., & Gallagher, S. (Eds.). (2018). *The Oxford handbook of 4E cognition*. Oxford University Press.

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège y J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 88–112). Oxford University Press.

Pérez, D., & Gomila, A. (2022). *Social cognition and the second person in human interaction*. Routledge.

Reddy, V. (2005). Before the 'third element': Understanding attention to self. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 85–109). Oxford University Press.

Reddy, V. (2008a). *How infants know minds*. Harvard University Press.

Reddy, V. (2008b). Experiencing others: A second-person approach to other-awareness. In U. Mueller, J. I. Carpendale, N. Budwig, & B. Sokol (Eds.), *Social life and social knowledge: Toward a process account of development* (pp. 123–144). Taylor & Francis Group.

Rivière, A. (1986/2003). Interacción precoz. Una perspectiva vygotskiana a partir de los esquemas de Piaget [Early interaction. A Vygotskian perspective from Piaget's schemas]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comp.) *Ángel Rivière. Obras Escogidas, Vol II* [Ángel Rivière. Collected Works, Vol, II] (pp. 109–142). Panamericana.

Rochat, P. (2001). *The infant's world*. Harvard University Press.

Rowlands, M. J. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. The MIT Press.

Schaffer, R. H. (1977). *Studies in mother-infant interaction: Proceedings of the Loch Lomond Symposium*. Academic Press.

Scotto, C. (2002). Interacción y atribución mental: la perspectiva de la segunda persona [Interaction and mental attribution: The second-person perspective]. *Análisis filosófico, 22*(2), 135–155.

Sheets-Johnstone, M. (2011). *The primacy of movement*. John Benjamins Publishing.

Sheets-Johnstone, M. (2015). *The corporeal turn: An interdisciplinary reader*. Andrews UK Limited.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist, 68*(8), 618–629. https://doi.org/10.1037/a0034185

Stern, D. (1985). *The interpersonal world of the infant. A view from Psychoanalysis and development psychology* [El mundo interpersonal del infante. Una perspectiva desde el psicoanálisis y la psicología evolutiva (J. Piatigorsky, Trad.). Buenos Aires: Paidós, 1991]. Basic Books, Inc.

Stern, D. (2000). Putting time back into our considerations of infant experience: A microdiachronic view. *Infant Mental Health Journal: Official Publication of The World Association for Infant Mental Health, 21*(1-2), 21–28.

Stern, D. (2010). *Forms of vitality: Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford University Press.

Thelen, E. (2000a). Grounded in the world: Developmental origins of the embodied mind. *Infancy, 1*(1), 3–28. https://doi.org/10.1207/S15327078IN0101_02

Thelen, E. (2000b). Motor development as foundation and future of developmental psychology. *International Journal of Behavioral Development, 24*, 385–397. https://doi.org/10.1080/016502500750037937

Thelen, E. (2005). Dynamic systems theory and the complexity of change. *Psychoanalytic Dialogues, 15*(2), 225–283. https://doi.org/10.1080/10481881509348831

Thelen, E., Corbetta, D., & y Spencer, J. P. (1996). The development of reaching during the first year: The role of movement speed. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1059–1076. https://doi.org/10.1037/0096-1523.22.5.1059

Thompson, E. (Ed.). (2001). *Between ourselves: Second person Issues in the study of consciousness*. Imprint Academica.

Thompson, E., & Varela, F. J. (2001). Radical embodiment: Neural dynamics and consciousness. *Trends in Cognitive Sciences, 5*(10), 418–425. https://doi.org/10.1016/S1364-6613(00)01750-2

Trevarthen, C. (1982). The primary motives for cooperative understanding. In G. Butterworth & P. Light (Eds.), *Social cognition* (pp. 77–109). Harverster.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). Cambridge University Press.
Trevarthen, C. (2017). The function of emotions in early infant communication and development. In *New perspectives in early communicative development* (pp. 48–81). Routledge.
Trevarthen, C., & Reddy, V. (2007). Consciousness in infants. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 41–57). Blackwell Publishing.
Varela, F. J., Rosch, E. y Thompson, E. (1991). *The embodied mind. Cognitive science and human experience*. MIT Press.
Valsiner, J. (2014). Needed for cultural psychology: Methodology in a new key. Culture & Psychology, 20(1), 3–30.

Silvia Español holds a PhD in Psychology from Universidad Autónoma de Madrid (Spain). She is currently a researcher at the Latin America Institute of Social Research (IICSAL), Faculty of Latin America Social Sciences/ National Scientific and Technical Research Council (FLACSO/ CONICET, Argentina). She coordinates the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc in Cognitive Psychology, and Director of FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is a certified Somatic Educator in the Feldenkrais Method. Her work deals with cognitive developmental psychology, psychology of music, and the study of human movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/silvia.espanol

Mauricio Martínez has a PhD in Psychology from Universidad Autónoma de Madrid (Spain). He participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning at the Faculty of Latin America Social Sciences (FLACSO). He is Professor of Developmental Psychology at Universidad Abierta Interamericana (UAI) and Faculty member of the MsC. in Cognitive Psychology (FLACSO). He also teaches in the post-graduate virtual course The baby's life: psychological development, scaffolding and nurturing environment" at FLACSO. His research focuses on early perceptual and intersubjective development, childhood play and early intervention in autism spectrum disorders. He is author of Intervención psicoeducativa para niños con trastorno del espectro autista [Psychoeducational intervention for children with autism spectrum disorders] (2015). His publications are available at: https://www.researchgate.net/profile/Mauricio-Martinez-2

Fernando G. Rodríguez is PhD in Educational Sciences from Universidad de Buenos Aires. He has carried out research and training stays in Germany (DAAD) in Linguistics and Semiotics. He is currently Professor of Semiology and Human Communication. He is member of the Research Group of Socio-Cognitive Processes in Infancy and Early Childhood at the Psychology of Knowledge and Learning Department (FLACSO, Argentina). As a researcher, his work is concerned with the development of communicative skills in young children, the structure of semiotic processes in general and the relationships between mind, sign and corporeality. Author of *Semántica y pragmática en la comunicación del niño pregramatical. Gesto y palabra antes de la articulación morfosintáctica* [Semantics and pragmatics in the communication of the pregrammatical child. Gesture and speech before morphosyntactic articulation] (2021) and co-editor of *Los signos del cuerpo. Enfoques multimodales de la mente y el lenguaje* [The signs of the body. Multimodal approaches to mind and language] (2018). His publications can be found at https://www.aacademica.org/fernando.gabriel.rodriguez

Chapter 2
The Infant-Directed Improvised Performances: What They Are and What Happens Through Them

Silvia Español, Favio Shifres, Isabel Martínez, and Diana Pérez

Abstract When adults interact with babies, they do special things. In this chapter, through a microanalysis of interaction scenes between an adult and a 7-month-old baby, we describe the infant-directed improvised performance. The aesthetic perspective assumed led us to continue the path initiated by others, evidencing remarkable structural and functional affinities between these kinds of encounters between adults and infants and the temporal arts. The infant-directed improvised performances are sound-kinetic phrases improvised by the adult through resources also used in temporal art performances, like the repetition-variation form. Adults create brief motifs collected from the baby's behavior or contingencies in the surroundings and repeat them in varied forms. In this chapter, we specify the temporal, energetic, and spatial dimensions of these variations through the use of analytical tools developed for the exegesis of artistic expression. Interesting events occur in infant-directed improvised performance: adults summon infants to social life, offer them well-formed behavioral units favoring their recognition, and interpret the world for infants, regulating their moods. They also illuminate cognitive structures such as image schemas and invite infants to experience primary metaphors. By virtue of performances, babies are also initiated into corporeal and aesthetic enculturation.

Keywords Infant-directed improvised performance · Primary metaphors · Image schemas · Temporal arts · Expressive resources · Improvisation

S. Español (✉)
Latin American Institute of Social Research (IICSAL), Latin American Faculty of Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina
e-mail: sespanol@flacso.org.ar

F. Shifres · I. Martínez
Laboratory for the Study of Musical Experience, Faculty of Fine Arts, Universidad Nacional de La Plata (LEEM-FBA-UNLP), La Plata, Argentina

D. Pérez
Philosophical Research Institute/SADAF-CONICET, Universidad de Buenos Aires, Buenos Aires, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_2

2.1 Introduction[1]

The term infant-directed performance usually refers to multimodal adult-directed information to infants. We incorporate the concept of *infant-directed improvised performances* to address a more restricted phenomenon bearing particular features that we revealed using microanalysis, pushing us beyond naturalistic observation. Our use of the term is restrictive but rests on its frequent application and previous theoretical advances in developmental psychology that made infant-related adult behavior and early interactions an appealing object of study. First, we will introduce the theoretical context that enabled the identification and assessment of infant-directed improvised performances. Then, we will discuss their form and function, focusing on the cognitive, affective, and aesthetic processes occurring through them.

During the twentieth century, language became the central research topic in the human sciences. It became involved in nearly all human affairs: in the construction of reality and subjectivity, our criteria of truth, and our interpersonal relationships. The so-called linguistic turn in the social sciences, along with the pragmatic and hermeneutic turns, emphasized intersubjectivity, revealing, among other things, the link between language and action. It shaped many diverse theories across the human sciences, at times barely compatible with each other, and was undoubtedly the conceptual birthplace of crucial studies in developmental psychology that underlie the concept of performance.

At the beginning of the 1960s, Austin (1962/1975) proposed that speech acts must include three synchronously performed acts – a perlocutionary act (pointing to the effect on the listener), an illocutionary act (involving the intention of the speaker), and a locutionary act (what is said). In the following decade, Bates (1976) suggested that the perlocutionary, illocutionary, and locutionary components could consecutively materialize in the course of development. The possibility of thinking about perlocutionary and illocutionary acts before the actual locution prompted psychological research. It then became apparent that the child's intention to communicate with others lies beneath the gestures and vocalizations taking place before speech. This now obvious idea is one of the main changes produced by the linguistic turn in developmental psychology. Not only did it change the way of understanding young children, but it also extended to the understanding of infants' early life in their social environment. Adult speech was found to change ostensibly when addressing infants, albeit not consciously or voluntarily, and to affect them, and maybe for the first time, nonverbal intersubjectivity was considered to be experienced early in life during adult–infant dyadic interactions.

From the 1970s onward, infant-directed speech and early intersubjectivity experiences became genuine research topics once the pioneering studies on early infant perception/action abilities fulfilled an adequate theoretical position: the effects of infant-directed speech are relevant for infant development only if we assume infants

[1] This chapter is based on content previously published in Español and Shifres (2015) and Martínez et al. (2018).

can perceive their features. We can only think of early intersubjectivity experiences if the infant is adult-oriented; if their subjectivity can be in touch with the adult's subjectivity. The infant assumed by William James, Jean Piaget, Sigmund Freud, and in some respects Lev Vygotsky was not a possible recipient of these speech transformations: if his perceptual discrimination and categorization capacities are null (if blind and deaf at birth), if he is self-oriented, solipsistic, or egocentric, and has no social skills or tendencies, then adult speech transformations either go unnoticed or are *affective nonsense*, and adult–infant intersubjective encounters are a vague parental illusion. But if, as cognitive psychology suggests, infants are oriented and attracted to people from the very beginning, if they perceive the sound of language when still in the womb and prefer it over other sound stimuli upon birth, and can recognize their native language and discern it from others, then we are in a position to assume they are affected by adult speech and are able to engage in psychological contact with others from the very beginning.

Psychological research found plenty of evidence of infant perceptual/action capabilities – among others, a preference toward contingent responses (Markova & Legerstee, 2006; Reddy et al., 2007), the detection of amodal temporal properties like duration or rhythm (Lewkowicz & Marcovitch, 2006; Bahrick & Hollich, 2008, see Chap. 3), and face recognition (Slater et al., 2010). Psychological research also focused on early intersubjective experiences grounding our interpersonal life. Around the early 1970s, Trevarthen incorporated the term intersubjectivity into developmental psychology (Bråten, 1998). Trevarthen (1982, 1998, 2017) elaborated the distinction between *primary intersubjectivity* – a dyadic psychological way of mutual contact developing between the ages of 2 and 9 months – and *secondary intersubjectivity* – whose unquestionable manifestation is the emergence of joint attention at around the age of 9 months. While the distinction suffered some criticism (Reddy, 2005, 2008), all subsequent psychological research on early interactions and social-cognitive development was marked by the idea of intersubjectivity.

Guided by *intuitive parenting* – a general and natural ability to protect, feed, stimulate, and teach cultural features to infants (Papoušek, 1996a; Papoušek & Papoušek, 2002), adults spontaneously act in a particular way toward infants, different from their acting with adults. Infant-directed speech (from now on IDS) is the paradigmatic example of this "special performance." A higher pitch, expanded pitch contours, rhythmicity, repetition, and reduced speaking rate make IDS sound much more musical than adult-directed speech (Fernald, 1989; Papoušek, 1996b). Thanks to their early perceptual abilities, infants can interpret these speech sound features (Trehub, 2010, 2015). For example, parents tend to use a set of five or six melodic prototypes with infants responding to their inherent messages. Adults use ascending contours or offer a turn in dialogue to gain attention and descending contours to withdraw or upon overexcitement. Extended bidirectional contours prevail when aiming to reinforce or approve some action. Parents tend to repeat their prototypical melodic contours but also vary them constantly recalling music organization standards: repetition-variation, tension, and arousal regulation (Fernald, 1989; Papoušek, 1996b). Such repetition-variation form is a flawless stimulus because (1) a constant

stimulus produces habituation and infants lose interest; and (2) repetition induces regularity, which helps to anticipate what comes next (Papoušek, 1996b; Rivière, 1986/2003; Stern, 1985). Studies on IDS have shown (a) its functions concerning the arousal and regulation of the infant's emotional states, favoring language acquisition, (b) its changes as an adaptation to infants' development throughout their first two years (Papoušek, 1996a, b).

Another radical effect of the linguistic turn in developmental psychology was the incorporation of the *proto-(...)* formula referring to any preverbal manifestation of a language-like event. Protodeclarative, protoimperative, protonarration, and the earliest in ontogenetic terms: protoconversation make up the list. They involve two people in a sign exchange situation where turn-taking is co-created (Bateson, 1979). Trevarthen repeatedly highlighted (1982, 1998, 2017) the pivotal role that Bateson's description of protoconversations played when he developed the concept of primary intersubjectivity. Protoconversations reflect an interpersonal form of cooperation underlying the development of later social capacities (see Chap. 3 for a more extensive description of protoconversations). More recently, Malloch and Trevarthen (2009) analyzed the pulse-regulated temporal organization and tone characteristics concerning the fundamental frequency and harmonic spectrum in mother and infant vocalizations during protoconversations. As we will later see, incorporating these items into the analytical framework led them to identify the fundamental features of communicative musicality. Turn-taking skills essential for speech, conversation, and adult dialogue (Levinson & Torreira, 2015) take their primary form in these mutually regulated adult–infant dyad exchanges that are still analyzed today (Gratier et al., 2015).

But speech and corporeality are not independent of each other. Since the 1970s, IDS' vocal elements are thought to be embedded in multimodal patterns of preverbal communication that include tactile, kinetic, and vestibular stimulation. Adults direct repeated and varied multimodal patterns to infants creating larger units of kinetic and vocal phrases. Kinetic and vocal phrases can run separately, jointly, or alternately (Stern, 1974; Stern et al., 1977). There is a repetition of smaller and discrete maternal-act units conforming to more extended ones according to the infant's response. Once the caregiver gains attention, the baby smiles, moves his head, or vocalizes, stimulating more concatenated maternal acts. This temporal organization of multimodal adult behavior emerges during the first months of life and recedes during the second half of the first year, once babies twist their interest toward surrounding events and objects (Stern, 1974; Stern et al., 1977).

The linguistic turn also gave prominence to the interactive synchrony between adult speech and some infant movements. According to Condon and Sanders (1974), from the first days, babies perform precise and sustained segmented movements synchronic with the articulatory structure of adult speech, resembling an *interactive dance*. Additionally, these authors noticed that infants and caregivers spontaneously engage in a sequence of contingent facial expressions and vocalizations to which infants are acutely sensitive, becoming disturbed when the dance is interrupted. When mothers stop responding and show a still face, infants fuss, cry, and eventually disengage (Tronick et al., 1978; Weinberg & Tronick, 1996). Conversely, when

mothers smile, exaggerate facial expressions, and bob their heads, infants are more likely to smile, vocalize, and laugh (Kaye & Fogel, 1980). Unfortunately, the initial extremely laborious coding analysis of body movements after watching videos frame by frame was set aside (maybe due to technical advances that facilitated the audio-computer analysis of adult–infant vocalizations). The pioneering works were rapidly abandoned and replaced by studies restricted to examining the temporal structure of dialogue (Jaffe et al., 2001).

The emergence of new technical advances restored movement (alongside sound) to its glorious past in developmental research. Using technological advances to capture and analyze movement, Brand et al. (2002), Koterba and Iverson (2009), and Licata et al. (2014) examined the movement of mothers when interacting with infants eight months and older from an action-to-goal perspective (like grasping an object), finding they modify them. They coined the term "motionese" to address the modification and simplification of gestures, actions, or signs used by adults while interacting with infants or toddlers. Motionese could specifically help infants to pay attention to objects, explore them, and process human actions. Kalashnikova et al. (2017) reported that phonatory movements (tongue and lip movements) articulated with IDS are independent of the specific sound features of this speech modality and that their mutability most likely responds to emotional variables. Focusing on the locomotor play of walking infants, Hoch et al. (2021) found that the infant–mother social dance continues once infants are free to move around the floor. As they gain new locomotor skills, early interactions expand to a complex spatial-temporal social dance when giving their first steps.

Through some theoretical constructs, we can interpret different aspects of the mother–infant *dyadic dance*. For example, mutuality and reciprocity have become global and meta-theoretical concepts involving dyadic processes during mother–infant interaction. They differ in how they underline the contributions of each participant of the dyadic system (Provenzi et al., 2018). Mutuality is often considered in studies emphasizing the bidirectionality and asymmetry of early exchanges: generally speaking, the caregiver is responsible for maintaining dyadic synchrony, while infants tend to drive the interaction through the caregiver's induction of contingent responses to their behaviors (Beebe et al., 2016). The term reciprocity predominates when exchanges seem symmetrical; for example, Trevarthen (2017) considers adult behavior part of a developmental strategy centered on the infant's intentional behavior with whom he co-regulates participation. In this chapter, reciprocity and mutuality are used undistinguishedly because the discussion exceeds its scope. Considered the soul of second-person encounters, reciprocity acquires diverse exchange modes. For Provenzi et al. (2018), dyad reciprocity (and mutuality) involves at least seven extensively examined processes: attunement, contingency, coordination, matching, mirroring, reparation, and synchrony, which are also present in infant-directed improvised performances.

2.2 About Infant-Directed Improvised Performances

Between infants and caregivers, there is a complex and varied dyadic dance featuring moments of harmony and disharmony. Infant-directed improvised performance is the name we gave to a specific kind of reciprocal interplay, an emergent harmonic organization in the complex, dynamic, and ever-changing dyadic interaction. It rests on the adult's ability to improvise based on the repetition and variation of simple motifs. It was first noticed in the context of a theoretical perspective rooted in Stern's work that intends to understand developmental processes in the light of artistic experience.

2.2.1 The Aesthetic Perspective

Stern's description of adult-offered multimodal patterned performances to infants (Stern, 1974; Stern et al., 1977) and the idea of an infant resembling a music listener or an abstract dance spectator (Stern, 1985, 2010) are the clearest antecedents of our conception of infant-directed improvised performances. The link between early reciprocity experiences and the arts became increasingly relevant in developmental studies. Dissanayake (2000, 2001, 2017) suggested the existence of temporal arts antecedents in early interactions. To attract attention, the caregiver elaborates multimodal performances, *making special* ordinary sounds and movements by repeating phrases, sounds, movements, and facial expressions and by exaggerating the melodic contours, amplitude, duration, and pauses in them. This elaboration – entailing saliency, novelty, expectation, and therefore emotion – is powered to lead directly to a state of inherently enjoyable mutuality. Dissanayake's concept of performance is extensive, close to Stern's description of multimodal packages.

Miall and Dissanayake (2003) were pioneers in the use of poetic resources while studying infant-directed vocal and kinetic performances that typically appear between the infant's sixth week and sixth month. They found parallelisms, hyperboles, alliterations, assonances, rhymes, and episode organization with a clear beginning or introduction, ending, and, sometimes, refrains or coda. Dissanayake (2008, 2017) suggests that *artification* is the ancestral activity or behavior that gave rise to, and continues to characterize, all instances of what today we call art. According to Imberty (2002), the repetition-variation form in adult–infant behavioral sequences represents the original structure whose profound reality music will only reactivate: with a sufficient margin of uncertainty, they both offer an invitation to remember and anticipate, insinuating the feeling that next time repetition might be left unaccomplished, that the future is unknown, and that the same expectation may merge into another that in turn may very well not be entirely different.

The performing arts involve artistic manifestations with expressions conditioned by their unfolding in real-time (we refer to musical performance and dancing, as opposed to musical scores or choreographies). Recently, two theoretical constructs

emerging at the intersection of early interaction and performing art studies provide resources to help understand early interactions while highlighting the role of movement.

The first one is *communicative musicality* (Malloch, 1999/2000; Malloch & Trevarthen, 2009). It entails interaction modes where both participants' rhythms, sound, and motor gestures match and coordinate, offering a particular way of being together from birth and throughout life, in communal activities and temporal or performing arts. Our biological and psychological structures involve a component that significantly contributes to a kind of communication that may be characterized as *musicality*. This component facilitates the establishment of a progression based on the regularity and predictability of social behaviors, out of which novelty and surprise originate. The term communicative musicality encompasses interactive, bodily, temporally organized, extensive phenomena associated with (and broader than) the cultural artifact of music. Communicative musicality can be understood as the general ability to match sound and kinetic gestures in terms of their spatial and temporal configuration. It depends upon the ability to coordinate them according to three main features. The first one is *the pulse*, which determines the way actions are regulated in time through the establishment of regularities. The temporal characteristics of actions and their occurrences get organized by temporal patterns whose uniformity makes them predictable. The second one is *quality.* It refers to how sound and movement are modeled in terms of their nontemporal attributes. The kinetic qualities are linked to the gesture's shape and contours, while the sound qualities to the pitch and spectrum of sound components. The third one, *narrative*, accounts for the organization of action sequences in successions acquiring meaning through their episodic configuration. While interacting with others, these three components foster encounter and coordination. Malloch and Trevarthen (2009) demonstrated that the play of a two-month-old with an affectionate mother shapes through the pulse, affective quality, and narrative expectation of musicality. The baby performs skillfully in the choreography of their improvised duet, taking turns to produce visible expressions and vocal sounds. Mother–infant interaction relies on timing mechanisms similar to those used by improvising musicians when adjusting each other's behavior (Malloch, 1999/2000; Stern, 2010). Mother and infant set up an improvisation zone of temporal structures where they play with time. Repetition implies appropriation and projection. The variation of repeated forms and processes transforms the known into the new, moves it forward, and invites it into an arena for creative dialogue and exploration. The variations in timing, intensity, and duration in a musical performance enhance aspects of the musical structure by facilitating the segmentation of sequences. In mother–infant interaction, variations of acknowledged forms may contribute to the strengthening of a shared proto-habitus (Gratier & Apter-Danon, 2009). As we have already suggested, the dual actions of anticipating familiar forms and processes, and perceiving and producing novel ones based on them, bring crucial dynamism and vitality to the ongoing interaction. This holds true for multiple forms of interaction: pre-verbal, conversational, musical, and kinetic.

Forms of Vitality is the second theoretical construct coming out from the intersection of early interaction and performing art studies. The term emerged within developmental psychology to account for those affective encounters between adults and infants that could not be described in terms of Darwinian emotions (anger, joy, sadness, fear). It was proposed by Stern (2010) to capture how the human mind deals with dynamic experiences that are crucial in interpersonal encounters and performing art experiences. Forms of vitality can be grasped from experiences and can be directly observed in the behavior of others. They relate to the way of doing things. A form of vitality is a *Gestalt*, a spontaneous integration emerging from the holistic experience of movement, time, force, space, and directionality/intentionality. We may bite, write, or draw in an eager and anxious or distracted and delayed way. Each way of performing any movement or vocalization involves a profile of activation. Vitality forms are activation/arousal profiles picked up directly, requiring no inference. They provide information about the affective states of the participants. In this sense, they are a primary way of relating to and understanding others and experiencing ourselves (Stern, 1985, 2010). The forms of vitality favor the transparency of people's subjective states. They are one of the means by which, from the beginning of our social life, we are not opaque to each other, but rather – as Reddy (2008) argues – have a fundamental perceptual union of active reciprocity. In a later work, Trevarthen (2017) picks up the concept of forms of vitality. He says that the playful interactions between adults and infants after the second month of life increase in liveliness, offering rich evidence of the baby's intuitive awareness of his partner's motives and how to trigger them. He considered that Stern's theory of vitality dynamics shows how we move with controlled elegance in the emotive common sense of an interpersonal world shared from the time when we were infants. And infants show and respond to the vitality dynamics in all acts of orientation and the expression of feelings for communication. For a baby, as it is for all social animals, play becomes the joyful language of cooperation and the essential way to take a role in the rhythms of a common sense of what the shared world affords.

Both phenomena – communicative musicality and forms of vitality – were observed during the first year: the first manifestations of communicative musicality tend to appear toward the second month, with a peak of increasing ability at about the sixth and seventh months (Trevarthen & Reddy, 2007); likewise, in early social play parents use movement and sound modulating the forms of vitality to enjoy intersubjective encounters (Stern, 2010, see Chap. 9). Both theoretical constructs relate early interactions and the temporal arts, emphasizing the continuity between the intuitive behavior of early intersubjective encounters and more elaborated ones involved in the social phenomena we recognize as artistic expression. According to them, the repetition-variation form is a core feature in IDS, music, dance, and poetry; therefore, it is also the core of infant-directed improvised performances.

2.2.2 *The Varied Repetition of Simple Motifs: A Strategy for Reciprocity*

Within the varied and complex array of infant-directed multimodal packages (Stern) or performative behavior (Dissanayake), there are some that we find peculiar, namely, those collecting that common adult strategy for reciprocity: the multimodal repetition-variation of simple motifs (Español, 2014). We call them "infant-directed improvised performances" (hereafter IDI-performances). We are interested in the multimodal packages containing simple motifs repeated at least twice. They can be fully and satisfactorily described with tools borrowed from two temporal arts: music and dance (the arts of sound and movement in time). To identify IDI-performances, we focus on what they repeat (the motif) and the dimensions of the aesthetic modeling of the repetitions (how they repeat); this has been the target of our studies. Given their frequent occurrence (see Chap. 3) and their striking aesthetic elaboration, they need to be identified and highlighted.

The norm in the dyadic dance is diversity, and IDI-performances become a fruitful, organized, harmonious moment of this process. Sometimes they are chained together, generating longer interactive episodes. Other times they bind with protoconversations or are interrupted by everyday life circumstances. Within the IDI-performances, adults guide the organization of events, but the baby's participation is essential for them to unfold. The infant responds to a call into social life, reveals the influence of adult IDI-performance, and provides reciprocity patterns to encourage her partner. He also regulates his IDI-performance (through smiles, kicks, gurgles, vocalizations, head movements), indicating he is pleased and wants to keep it going or that he has had enough and would like it to end. The caregiver provides multimodal information – sound, visual, kinetic, and tactile – simultaneously available to the different perceptual systems. The infant perceives this complex information differently from the adult. Identifying the features of what they perceive, determining to what degree and how they vary throughout development are questions bearing some answers and awaiting others (see Chaps. 6 and 7 for these issues).

The infant inspires adult participation, offering what soon could become a motif for improvisation (if he sneezes, bites, smiles or vocalizes, or suddenly shows his teeth, moves his arms, legs, or head, the adult selects the sound and/or movement made by the infant, repeats and transforms it). Motifs are also extracted from the surroundings (a noise, a change in lighting, something that moves or suddenly appears) or from the adult (bodily sensations or memories). Motifs are small sound-kinetic units usually lasting between 2 and 5 seconds. With repetitions, they become short phrases lasting between 10 and 50 seconds, approximately. Together, motif and phrase constitute the IDI-performance (Español, 2004).

Unlike the phenomenon known as infant-directed speech, IDI-performances are essentially based on movement and may not include speech. While the continuous stream of movement knows of partial pauses and constantly disseminates throughout our body, speech is intermittent.

The motif can be a word or a sequence of words but, most frequently, it is a nonverbal gesture-like behavior, a simple action like pulling out a blanket, kissing, touching, or movements on the baby's body, moving an arm, or holding the infant up in the air, or moving the legs rhythmically with the infant on the lap. Some repetitions of the verbal motif may omit speech, accompany, embellish, or make it more attractive by using nonverbal behaviors; conversely, nonverbal motifs may include speech in the repetitions. Speech emerges exhibiting traits of primary orality (Español, 2010). The variations that animatedly transform the motifs are very diverse. Most of them tend to brim with musicality. Adults improvise their repetitions in real time with grace and smoothness with intuitive and unconscious ability, molding sounds and movements, making them more striking and appealing to the infant. IDI-performances are shaped by motif reiteration, extending in time a little longer than other adult behaviors such as affect attunement or simple imitation (Español, 2014).

The aesthetic perspective of dyadic interaction reveals remarkable structural and functional affinities between early intersubjective contact and the communicative-affective experience of temporal art. This evidence led us to explore the IDI-performances with analytical resources used to study music and dance. In several works, we described the aesthetic modeling of IDI-performances by analyzing their motifs and their sound/kinetic variations (Español, 2014; Español & Shifres, 2015; Martínez, 2014; Shifres, 2014). We observed they seem to be permeated by musical patterns belonging to the culture of the adult. For example, the way they organize around a simple motif suggests an embryonic reflection of what will later constitute more sophisticated elaborated forms of artistic expression in the Western musical tradition, where compositions get organized around motifs as basic units or ideas (Martínez, 2014). Likewise, when repeating motifs, adults employ Western culture's expressive resources. For example, they use timing (*rubato*) and sonority gradients (strong-weak) for phrase closure (Shifres, 2014). Our intention in this chapter is to demonstrate how tools that come from music performance studies and movement analysis in contemporary dance allow us to link the structural features of sound and movement organization with the directly perceived meaning in a sound and kinetic adult composition, as it happens in the temporal arts.

2.2.3 Two Performing Art Tools to Analyze IDI-Performances

Music Performance Studies That Are Relevant to the Analysis of Multimodal IDI-Performance

Music performance studies emphasize music interpretation and communication between musicians and listeners (Gabrielsson, 2003; Juslin, 2013). These works aimed not only to describe music performance's expressive features but also to explain the intersubjective relations elaborated from them. The perspective of music performance studies connects the descriptions of sound musical phenomena with

characteristics that make each performance a unique and unrepeatable event. This can be achieved mainly by focusing on a behavioral pattern recognition that could explain idiosyncratic, intentional, and expressive performance aspects elusive to music notation. In this sense, it is not only important to recognize how the durations of sounds can be identified with durational discrete categories (proportional) but also how the actual sound duration systematically differs from those categories (Shifres, 2014). Defining systematic differences is crucial to account for action intentionality. The most important expressive variables are timing and dynamics. For their analysis, it is compelling (i) to see how the regulatory frame (the metrical structure and the dynamic range) sets; (ii) where that regulatory frame deviates (by lengthening or shortening periods or by increasing or reducing dynamic ranges); and (iii) what motivates such deviations (structure, emotion, communication, etc.). Concerning the timing, performance studies sought to find behavioral patterns related to expressive timing according to structural aspects of the composition. Similarly, other aspects of expressive characterization of sound in music, such as dynamics (or loudness modulation) and articulations (the expressive characteristic derived from the length of the pauses between two consecutive sounds, combined with the type of the sound's attack), are relevant. Finally, it is worth mentioning that music performance studies progressively covered other areas beyond structural explanations. Since the 1990s, more than a few studies have inquired about the expressive deviation related to the intended emotional content (Gabrielsson & Lindström, 2010) or kinetic content (Kendall & Carterette, 1990).

Performance studies highlight that the expressive microstructure (the set of systematic features of timing, dynamics, articulations, etc.) carries information interpreted by the listener in emotional, grammatical, bodily, and narrative terms. The ability to detect such messages in microstructural systematizing correlates with the proficient recognition of the internal state of a partner in reciprocity experiences (Shifres, 2014, 2020). A complete description of the reciprocity phenomenon must characterize its sound elements in terms of music categorical features – such as musical tones, meter, proportional durations – and noncategorical aspects – such as pitch quality (Malloch, 1999/2000) and microstructure relationships of timing, dynamics, and articulations (durations) (Shifres, 2014).

Contemporary Dance Studies Relevant to the Analysis of Multimodal IDI-Performance

Movement analysis in early interactions is far less sophisticated than that of sound: in general, it stresses the moving parts of the body, their synchronous relationship to speech, the exaggeration of some gestures and facial expressions, and their concordance with other aspects of adult multimodal stimulation. The dynamic experience of movement became a focus of attention with the emergence of modern dance. According to Laban (1971), each movement phrase, every little weight transfer, or the simple gesture of a body part reveals a feeling trait. Moves can be performed with multiple expressive variations: apples can be eaten hungrily or nervously, with

focused movements triggered rapidly, suddenly, or in a carefree way with scattered soft and slow movements that seem incomplete. When choreographers looked into nonlinear, random, and nonnarrative forms, the expression of movement came to the fore (Sheets-Johnstone, 2011, 2015; Stern, 2010).

Laban-Bartenieff's Movement Analysis System describes movements from their *shape* and *effort* (Davies, 2006; Laban, 1971; Moore, 2009; Newlove, 2007). Shape and effort capture the dynamics and expression. Shape comprises (i) body: it concerns body parts used in the observed movement; (ii) dimension or space: it concerns the relation of the body with the surrounding space, with the kinesphere being a space limited by the points reached by the limbs when keeping the same body position and defining three body "attitudes" in this space: vertical (up-down), horizontal (right-left), and sagittal (front-back); (iii) form: understood in terms of the basic opening/closing opposition when breathing, on the vertical, horizontal, and sagittal planes, obtaining three forms: rising-descending, spreading-enclosing, and advancing-retiring. Effort indicates the quality of each movement based on a dedication or fighting attitude toward four motion factors: space (direct vs. flexible), time (sudden vs. sustained), weight (strong vs. light), and flow (free vs. bound). While combining the first three factors, eight basic types of effort emerge: (1) pressing (direct-sustained-strong), (2) gliding (direct-sustained-light), (3) punching (direct-sudden-strong), (4) dabbing (direct-sudden-light), (5) wringing (flexible-sustained-strong), (6) floating (flexible-sustained-light), (7) flicking (flexible-sudden-light), and (8) slashing (flexible-sudden-strong). They can all be performed in a continuum flow (the fourth factor): from a free flow (difficult to stop) to a constrained flow (easy to stop).

The theory highlights the feelings running along with these efforts. Every basic type of effort is accompanied by a movement sensation: pressing by a sinking or a heavy-long-threadlike sensation; gliding by a light-long-threadlike sensation of elatedness; punching by a heavy, short-threadlike dropping sensation; dabbing by a light-short-threadlike stimulation sensation; wringing by a heavy-long-pliant relaxation sensation; floating by a light-long-pliant suspension sensation; flicking by a light-short-pliant excitement sensation, slashing by a heavy-short-pliant collapsing sensation. The flow's movement sensation ties in with fluency: a free flow with a flowing on sensation, a leaping flow with the feeling of pausing or being stopped (Newlove, 2007).

The movement begins inwardly in the dynamosphere. It is all about the inner world, a psychological space, where impulses continually surge and seek an outlet through deeds, acting, or dancing. Laban thought the effort was a visible expression of thought and feeling. Attention, intention, and decision are stages of an inner readiness for external bodily action appearing once the effort can be concretely expressed through the body. The inner world of the dynamosphere encompasses many contrasting landscapes. Voluntary human movement springs from an inner intent turning into observable action modifying its quality, like adverbs modify verbs. A simple sequence of actions – such as entering a room – can be performed furiously, or slowly and casually. Contrasts suggest the performers´ internal states are mixed. Inner impulses cast dynamic shadows coloring the path of observable

movements. The kinesphere endures the physical geography of changing forms; the dynamosphere is a psychological landscape of shifting moods. Each body movement involves position changes in space using kinetic energy. Thus, the kinesphere and the dynamosphere penetrate each other. The interdependence of dynamospheric and kinespheric sequences reveals that the conventional idea of space segregated from time, force, and expression is completely erroneous (Laban, 1971; Moore, 2009). Movement and mind, body and soul, and emotion and concept are an entwined, inseparable wholeness (Lansdale, 2009).

Kinetic experience comes from many channels. Movement is seen and picked up proprioceptively with the participation of touch and the vestibular sense. Moving and experiencing a state of mind is the same thing. Laban's analytical approach virtuously grasps through visible movement (subject to description and analysis) the complex bodily and phenomenological movement experience. In this way, Laban-Bartenieff's movement analysis bridges the gap between moving from an objective observation to an interpretation in terms of subjective and intersubjective experience (Moore, 2009). Dance spectators are externally privileged movement observers. The project *Watching Dance: Kinesthetic Empathy* (2008–2011) used audience research and neuroscience to examine how viewers respond to and identify with dance. They explored the concept of kinesthetic empathy occurring when dance spectators sitting still feel they participate in the movements they watch, experiencing related feelings and ideas. The term kinaesthetic empathy expresses the "shifting, pulsating, writing, dancing, expressive action of bodies in space over time" (Jones, 2012, p. 12). It should be considered in terms of affect, with affect being embodied and preceding the cognitive differentiation separating our emotions into distinct and identifiable categories (happy, sad, etc.). Affect increases energy levels, and kinaesthetic affect relates to the impulse toward or the anticipation of movement rather than the actual movement (Reynolds, 2012).[2]

Considering effort as a visible expression of the inner world of feelings and thought is very similar (an antecedent) to the concept of forms of vitality. Terms like effort or mood closely relate to vitality forms in movement. In the words of Stern, "Laban was sensitive to what he called 'the spiritually vitalizing effect' of dance – of rendering movements that capture inner feelings – and the need for, and code different, even individual, interpretations (in short, the vitality forms of the dance). To capture the uniqueness of an interpretation he needed written dynamic indicators" (Stern, 2010, p. 85). On the other hand, the idea that kinaesthetic empathy

[2] From a first-person perspective in social cognition (focused on embodied simulation), mirror neurons and mirror related mechanism (a neural system matching action perception and execution of simple actions, gestures, movement, communicative mouth movement) were proposed as the neurobiological grounding for some forms of primary and secondary intersubjectivity (Ferrari & Gallese, 2007). Mirror neurons and other embodied mechanisms are considered to support the aesthetic experience in dance (Calvo-Merino, 2010), and visual arts (Freedberg & Gallese, 2007; Ferrari & Gallese, 2007). These findings endorse the idea that we are oriented or impelled to sympathetically feel certain movement qualities (or movement traces) of other people.

should be thought of in terms of affect (Reynolds, 2012) brings the concept closer to forms of vitality (or vitality affects) and their direct grasp (Nudler et al., 2020).

2.2.4 A Clear-Cut Description of IDI-Performances

When we analyzed adult-to-infant multimodal packages with the tools described above, we identified and defined IDI-performances as short multimodal phrases spontaneously constructed by adults for their babies based on the repetition-variation form of minor units or sound-kinetic motifs, modeled with expression. The temporal arts' categories of analysis helped to specify the multimodal motif modeling form. After reviewing and discussing our earlier formulations (Español, 2014; Español & Shifres, 2015; Martínez, 2014; Shifres, 2014), we propose that sound and movement modeling is spatial, energetic, and temporal.

Spatial modeling includes:

- For sound: Melodic contours and timbre variations.
- For movement: Shape variations, as described in the Laban system of movement analysis: rising-descending, spreading-enclosing, and advancing-retiring.

Energy modeling comprises:

- For sound: Degrees of relative sound intensity (*pianissimo* [very soft], *piano* [soft], *mezzopiano* [medium soft], *mezzoforte* [medium loud], *forte* [loud], *fortissimo* [very loud]).
- For movement: Effort variation, as described in the Laban system of movement analysis: (1) pressing (direct-sustained-strong), (2) gliding (direct-sustained-light), (3) punching (direct-sudden-strong), (4) dabbing (direct-sudden-light), (5) wringing (flexible-sustained-strong), (6) floating (flexible-sustained-light), (7) flicking (flexible-sudden-light), and (8) slashing (flexible-sudden-strong). Each of the eight basic types of effort can be performed in a continuum flow (the fourth factor): from a free flow (difficult to stop) to a bound flow (easy to stop).

Temporal modeling comprises:

- For sound and movement:
  - Speed variations: *ritardando, accelerando, rubato*.
  - Tempo variations: *largo, grave, lento, adagio, andante, moderato, allegro, vivace, presto, prestissimo.*
  - Behavior organization according to amodal temporal information: duration, rate (chronometric density), and rhythm.

2.2.5 *Between Affect and Cognition*

The activity performed by adults when facing an infant has been mainly studied from the affective regulative function perspective. However, in recent analyses of IDI-performances, it also appears to involve some intertwining between cognitive and affective processing. This can be seen when analyzing two preverbal and basic structures embedded in them: image-schemas and primary metaphors. They are included in the improvisation of the sound-kinetic IDI-performances that the adults deploy actively to-and-for the infants during intersubjective encounters (Martínez et al., 2018).

According to the conceptual metaphor theory (Johnson, 2007; Lakoff & Johnson, 1980/2003, 1999), image-schemas are dynamic patterns that emerge as a result of the sensorimotor activity of our recurrent body movements in the surrounding space. Image-schemas present a simple structure with very few elements. For example, the VERTICALITY[3] image-schema involves the recurrent experience of upward and/or downward motion that links two spatial locations. The SOURCE-PATH-GOAL image-schema is an abstract structure resulting from the enaction of the purposeful movement originated in a spatial location that, after deploying a linear trajectory, reaches a final destination. As far as we know, systematic observations of the links between primary metaphors and image-schemas in early interactions have not been pursued from this perspective. Moreover, the question regarding the evolutionary origin of image-schemas has always been made from an individualistic point of view on development. It is assumed that image-schemas emerge during the infant's sensory-motor activity (ascending, descending, standing up, and so on) without paying attention to the role of the caregiver in these experiences (Johnson, 1987; Mandler, 2005). Given that before moving by themselves in the surrounding space, infants are frequently moved by the adults, and adults move themselves in front of them, image-schemas somehow appear embedded in adult–infant interactions, becoming an important component of IDI-performances. Moreover, two image-schemas may appear combined during IDI-performances (i.e., the SOURCE-PATH-GOAL image-schema can be anchored in the VERTICALITY image-schema when the adult performs upward and downward movements in front of the baby) (Martínez et al., 2018). As to primary metaphors, Lakoff and Johnson (1999) also include them at the foundation of subsequent cognitive meaningful structures. Primary metaphors are basic structures originated by an overlapping or co-activation of neural patterns that will be conceptually distinguished later in development, but that remain bound at the neuronal level (Lakoff, 2008, pp. 26–28, gathering up the ideas of Johnson, Narayanan, and Grady). For example, in an early developmental stage, the primary metaphor AFFECTION IS WARM co-activates the subjective-emotional state of affection with the sense of temperature, conflating the state of feeling warm while being held/embraced with affection. Martínez et al. (2018) also

[3] For all cases of conceptual metaphors, the standardized use of capital letters in cognitive linguistics is respected.

suggested that some of the most basic primary metaphors that underline the affective components of experience would be somehow embedded at the core of such intersubjective adult–infant encounters.

Therefore, the experience of IDI-performances opens a path to gestaltic experiences of image-schemas and primary metaphors in inter-enactive contexts that will feed future individual experiences. As previously stated in this chapter, the structure of IDI-performances contains minor units or motifs – formed by rhythmic patterns, shapes of body movements, and/or vocal melodic contours – that are repeated slightly differently, including changes in spatial, energetic, and temporal dimensions. Martínez et al. (2018) run multimodal microanalyses of IDI-performances using a multimodal model (Martínez, 2014) to identify and characterize the allocation and features of the image-schematic forms and primary metaphors. The Multimodal Observational Model (Martínez, 2014) consists of three main components: (i) the sequence of actions of an image-schematic form performed following the classic canonic analysis of the image-schematic structure of the actions involved (Johnson, 1987, 2007; Lakoff & Johnson, 1999); (ii) the adult–infant expressive sound-linguistic components through which an image-schematic form and/or a primary metaphor is composed, enacted, and activated; and (iii) the type and quality (shape and effort) of the movements involved in the enaction of the image-schematic and/or a primary metaphor form. Martínez et al. (2018) show how image-schemas and primary metaphors are embedded and unfolded in those sound-kinetic phrases.

2.3 What Microanalysis Tells Us About IDI-Performances

We are now in a position to show (i) how adults improvise IDI-performances before babies through that simple strategy for reciprocity that is the varied repetition of simple motifs, (ii) how they vary their motifs by employing resources similar to those of artists in the performing arts, (iii) how IDI-performances link to other multimodal packages, and (d) how primary image-schemas and metaphors are brought to life within the IDI-performances.

In Sect. 2.3.1, to show (i), (ii), and (iii), we will partially reproduce some microanalyses and interpretations of an extended interaction scene analyzed in an article by Español and Shifres (2015). In Sect. 3.2, to show (iv), we will partially reproduce the microanalysis and interpretations of an interaction scene analyzed in Martínez et al. (2018).

2.3.1 The Repetition-Variations Form and Other Resources of the Performing Arts

Español and Shifres (2015) selected one adult–infant interaction scene from a previously recorded longitudinal case study research about preverbal communication. The study carried out by the first author of this paper featured interactions between the author and the first son of a Spanish middle-class family (Español, 2004). The participants were native Spanish speakers. The selected scene took place during the first interaction session between the researcher, Silvia (S), and the infant Habib (H); it was recorded at the baby's home when he was 7 months old. It is spontaneous, playful, and natural, showing everyday life contingencies. The researcher–infant dyad allows the experimenter to keep an online record of her own experiences and reactions. Thanks to the availability of recording audiovisual resources, she can later observe them remotely, retrieving her own previous experience. The selection of a researcher/infant dyad is consistent with Reddy's (2008) methodological proposal for studying intersubjective phenomena. Español and Shifres (2015) focused on adult IDI-performance, concentrating on sound and movement. The software PRAAT was used to analyze the sound and Laban-Bartenieff's Movement Analysis System to analyze the movement. We used the movement categories of shape and effort using the video annotation software Anvil 4.0 (Kipp, 2004). For this chapter, we selected two moments. At times, we will exhibit figures of the sound and/or movement microanalysis to show the tools used and to highlight some aspects of the sound and/or movement organization; in other cases, we will only make verbal remarks. For further details on the procedure and to access the complete microanalysis of the scene, see Español and Shifres (2015).

Brief Description of the General Scene

The baby is sitting on his baby seat; the adult is on the floor beside him. The 1.20-minute scene starts when the adult makes a movement to attract the infant's attention and ends when the latter turns his face around, probably to reduce his arousal level. It is a clear case of mutual behavior regulation. In between, we find a sequence of sound and movement units and moments well and poorly organized. We divided the general scene into Moment 1 and Moment 2.

Moment 1 – An IDI-Performance, a Phone Ring, and a Different Multimodal Pattern

Brief Description

The first section involves covering and uncovering the baby's body (trunk and legs) with a shawl. Initially, the infant (H) faces the opposite way while the adult (S) looks at him, making small movements to catch his sight unsuccessfully. Then she

slowly covers and uncovers him with the shawl to attract his attention. Immediately H starts to follow the direction of the cloth by turning his head until his gaze crosses that of S. From there on, what began as an almost casual movement becomes an ostensibly elaborated one, initiating a pleasant and joyful shared play. Moving the shawl in that manner is of paramount significance concerning the following kinetic-sound structure: first, it becomes the motif that will be repeated with variations, determining the thematic content of the IDI-performance as a whole, and second, it defines the elemental time span for the units that will structure the entire IDI-performance. After a silent beginning, the sound follows S's laughs (toward the end of the first repetition). Upon closure, a ring tone breaks the atmosphere. Moment 1 ends when S quietly returns to the center of her kinesphere following the interruption.

Multimodal (Movement and Sound) Microanalysis

Figure 2.1 shows that the varied repetition of the shawl removal action generates a movement phrase (23 seconds of duration). Repeated labeled rectangles exhibit the categories of the Laban Movement Analysis System that characterize the motif and its varied repetitions (elaborations) of a movement phrase. The microanalysis shows (i) the initial movement of removing the shawl, with a *gliding* effort (row 4), and a *retiring* shape (row 3); (ii) the elaborated repetition of the previous movement: a *gliding-floating* pattern in terms of effort, and a *rising-advancing-descending-retiring* pattern in terms of shape, with flow (row 5) alternating between *bound* and *free*; (iii) two other motif repetitions with variations in the effort pattern (*retiring-rising-descending and advancing-rising-descending*) and a variable extent of *retiring*; (iv) coda or closure with simplified movements by (a) taking away the object – the characteristic element of both the motif and its elaborated repetitions (row 1 alternating object-trunk labels) – and (b) simplification of the *rising-descending* pattern into the simpler *advancing-retiring* (row 3). The last advancing movement is the widest, marking the closest physical contact point between the dyad members and the most extensive movement (until reaching the target). The recurrent *retiring* movement closes the phrase. The adult withdraws and reaches the center of her kinesphere.

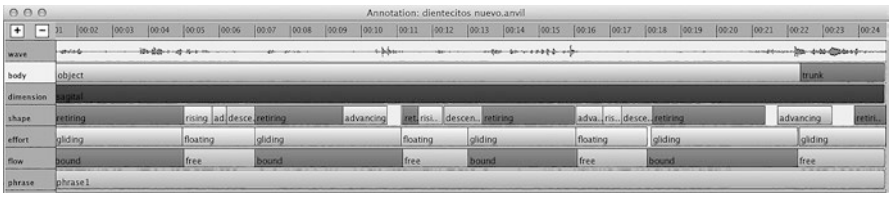


Fig. 2.1 Movement description of initial IDI-performance in Moment 1 according to the categories of the Laban-Bertenieff analysis using ANVIL 4-0. (From Español & Shifres, 2015). The horizontal numerical scale (above) indicates the seconds. The first row shows the dynamic envelope of the sound signal. The following rows correspond to the variables of the Laban system (body, dimension, shape, effort, and flow). Finally, in the last row are the units emerging from the analysis of these variables. (Reproduced from Español & Shifres, 2015)

The movement along the three repetition-elaboration events is temporarily regular. However, in the second repetition, the motif of retiring the shawl is accelerated, lasting one second less. This acceleration can be interpreted as expressive as long as (i) it is linked to H's demand to remove the shawl and S's caring attitude to meet that demand; (ii) it is the second repetition, where the expected time interval (the normative measurement of that interval) is already established, and the change is shown as a deviation from the expected measurement. Furthermore, the acceleration is joined by vocal sounds. As anticipated, S laughs at the end of the first motif repetition, enabling the sound dimension of the IDI-performance. During the second repetition, sounds are progressively shorter and higher. The repetition ends with a change in the voice register ("Ti-ti-ti-ti…"). In that way, she accompanies the shawl's removal in the second repetition with sharp, regular, short sounds (staccatos): "Ti-ti-ti-ti…." A temporal shortening configures in the multimodal unit, and the sound accompanies the progressive temporal movement reduction.

The third motif repetition extends, closing the entire sequence (last elaboration) like a final *ritardando*. This is understood as a temporal deviation not only for its closing location in the sequence but also for the way sounds run along with it: (a) first, the movement occurs in silence; soon the sounds expectation for the shawl removal increases, and (b) a breathed-in vocalization accompanies the last shawl withdrawal ("I take it out, I take it out, I take it out") accordingly, loudness decreases (diminuendo), and the sequence ends with an extended (*ritardando*).

The ending sequence of the movement elaboration follows – the simple advancing-retiring shape pattern – with a vocalization ("Uuuúpateee") taking place precisely upon advancing. It is the most extensive advancing movement of all repetitions, having the same duration as the "Uuuúpateee" sound. A silent recurrent retiring movement closes the phrase. The adult backs down. The "Uuuúpateee" vocalization is higher in pitch than the immediately previous one (similar to the "Ti-ti-ti-ti"). Returning to this *tessitura* rebuilds expectations. The generated anticipation suggests reaching a consistent unit that will satisfy expectations and close the loop. Although movement seems to close the previous episode, the sound tells the infant that the IDI-performance continues. Vocal sounds articulate in a *legato* style. Timing analysis of sound components allows appreciating the regularity of vocalizations and temporal relations among the phrases. Sound introduces a beat level of time regularity, and then a subordinate level (beat division). After reaching time consistency, the regular phrase sequence finishes extending the last beats of the last sound phrase.

The motif and the three repetitions make up an IDI-performance fully and easily improvised by the adult. Upon its end, S backs and reaches the center of her kinesphere. She takes the shawl up, covering her face and upper body exactly when the phone rings. The sound interferes with the intimate atmosphere of the performative plot progression just when what is going on seems to depend on a sound level thread, mainly because the pitch range and the loudness highly differ from those surrounding the scene. S lowers the shawl a little bit to uncover her face, pronouncing a breathy "Ahh." She freezes her movement and talks to the baby establishing eye contact. Here, freezing is a considerable movement innovation. Then she pulls

down the shawl and looks to the phone, looks back to the baby, and sustaining eye contact gets closer, with a soft movement at the sagittal plane, and says "*Te-lé-fo-no* (phone)" while returning to her kinesphere. The freezing responds to the ring tone and accompanies the expression of surprise, but then it breaks into the initial unit movement's repetition. The surprise and the abandonment of the movements' design appear to be offset by a systematic action at the sound level to regain control of the situation. From the first surprise sign, "Aaah!" a series of utterances follow one another ("Uy! Ah! Uh!"), retaking the initial pulse regularity. Similarly, these interjections retake the upper register. Notably, control is regained after 4 beats. In a low register to reduce tension and provide a safety message ending the interruption, S clearly pronounces with a syllabic isochrony the word phone: "*Te-lé-fo-no*." Moment 1 ends when they retrieve the confidence to continue exchanging pleasantly and joyfully a sort of peek-a-boo play.

The multimodal adult display of resources when the phone rings is full of musicality and kinesthetic novelty; however, it is not an IDI-performance but simply another adult multimodal package because it lacks the repetition-variation of simple motifs organization.

Moment 2 – A Peek-a-Boo Game, a Fortuitous Event, and a Resolution

Brief Description

Once the mood is restored, S resumes the interrupted peek-a-boo with the blanket. From behind the shawl, she says, "Habibi … Habibi … (appearing) Here is Habibi! (behind the shawl again) Habibi, I can't see you … Habibi, where are you? (appearing) Here you are!" Then a new fortuitous event alters the atmosphere: The back lock of the baby seat loosens, descending two positions. Immediately, the adult improvises handling the contingency that threatens the encounter's continuity.

Multimodal (Movement and Sound) Microanalysis

Generally, the peek-a-boo play presents two contrasting situations: one is losing eye contact (usually by the interposition of a blanket or another object), and the other is regaining visual contact upon the object removal. Unlike the initially multimodal IDI-performance in Moment 1, this one has a more complex, hierarchical, and embedded structure, probably due to its more tradition-bound nature. The movement analysis evidences an organization of kinetic features highly consistent with that of sound. The hierarchy observed in the sound analysis – the basic idea, the contrasting idea, and the antecedent-consequent phrase – may be discerned from the movement. When the consequent ends, the baby takes the initiative by catching the shawl and bringing it to the mouth. S says "Nooooo" with a descending melodic contour (like a typical categorical prohibition message) and a guttural vocal timbre and initiates the final retiring movement with a different effort (pressing) rising the coda: the game is over.

Compared to the multimodal IDI-performance in Moment 1, this one is not pure improvisation based on the repetition of a simple motif. We consider this a

performance because of the spatial and temporal modeling of sound and movement (see the details in Español & Shifres, 2015) and the repetition-variation. But what is repeated in this case is not a simple motif in a sequential manner; the IDI-performance adopts a more complex hierarchical form, composing a product based on the cultural repertoire of play. We call this a "ritualized infant-directed performance" to differentiate it from those *fully improvised that use the varied repetition of simple motifs.*

After the peek-a-boo's coda, a new fortuitous event alters the shared atmosphere: the back of the baby seat gives in, descending two positions. The sound component acquires a clear rhythmic-metrical structure that differs markedly from everything that was happening before. Consequently, the most noticeable feature of this fragment is the rhythmic structure. The speech becomes onomatopoeic, with marked consonants and shorter syllable durations.

Figure 2.2 shows the duration of each syllable and the pauses between them. Here, the *staccato* nature of the first sentence can be observed in the stressed accents and in the pulse regularity (beat). By contrast, in the second sentence, S lengthens the syllables, putting them closer together. She extends the syllable "Ay!" to a point where the emotional content changes. This point represents the division of the unit into two sub-units. The content of the first unit is surprise. The second one is the expression of empathy toward the disturbing situation suffered by the baby. As part of that empathy, when the next beat is due, she concludes with a low tone "*Esa ca* ..." (that she will pick up in the following IDI-performance – which will not be analyzed in Fig. 2.2 – completing "*Esa carita*"/that little face.../), touched by the infant's afflicted face. Figure 2.2 also shows differences between *sung* syllables – like "Uyyy" – and noisy syllables – like "Toc-toc." The former emphasizes the F0, while the latter exhibits more components in their spectrum.

This interruption differs from that in Moment 1 (a sudden phone ring) because it is multimodal: H's position changes (descending abruptly), and the two-time chair descending goes with two loud sounds coming from the unhooking lock. The strategy to overcome the situation is also multimodal: both stimuli (kinetic and sound) give S the motif to build the next behavioral unit and handle the contingency that threatens the encounter's continuity. The seat fell on the vertical plane with a direct, sudden, strong effort (punching) accentuated at the bottom. As Fig. 2.3 displays, S interrupts her movement in that instant, responding rapidly, imitating the fall with a similar movement but changing one element: moving the shawl twice with a light, direct, and sudden effort (dubbing), conserving the vertical plane. She quickly elaborates the unit by throwing her arms back: with a gliding and floating effort, moving them forward with two new and very smooth dubbing (like a smooth rocking) movements. The phrase closes with a freeze, reaching a state of calm, achieving the goal of the previous movement, and sustaining eye contact to make sure the baby regains ease.

Here a fortuitous sound/kinetic event gives S the material for her IDI-performance. The sudden chair fall experienced by the baby is a vertical, descending, and punching (direct-sudden-strong) movement conveying a collapsing feeling. In light of this event, S responds with a very similar but subtly different-movement: using the

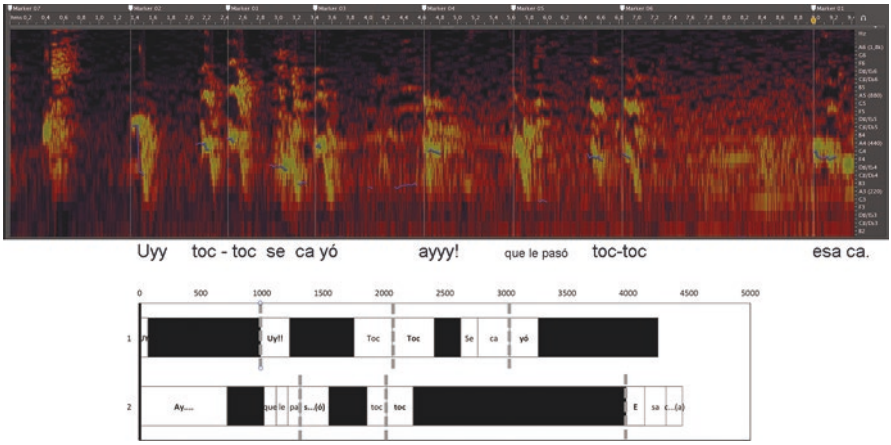


Fig. 2.2 Sound analysis of a Fortuitous Event in Moment 2 using PRAAT. (From Español & Shifres, 2015). Upper panel: spectrogram of the excerpt. The differences between vocalizations are shown according to their emphasis on F0 ("Uyyy" and "Esa ca…" are sung). The dotted vertical white lines show a regularity between the accented syllables (pulse). Lower panel: relative durations of sounds (white) and pauses (black) of a fortuitous event and resolution. The stressed syllables are marked with vertical black lines showing the pulse's regularity. The dotted vertical lines indicate a beat in the pause. Relations sound/silence account for the staccato or legato articulation of sounds and phrasing (way of articulating phrases using timing resources). (Reproduced from Español & Shifres, 2015)

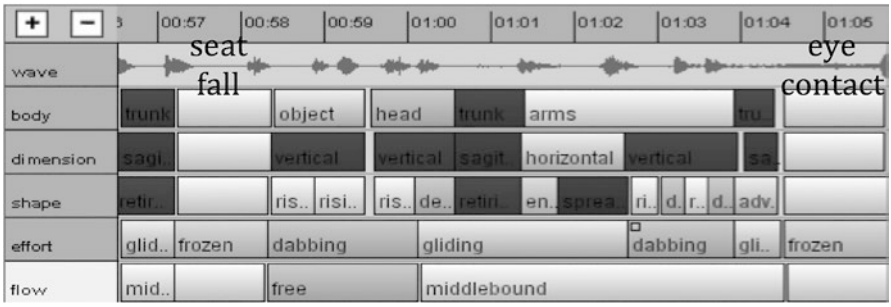


Fig. 2.3 The movement description of a fortuitous event and the resolution according to the categories of the Laban-Bertenieff analysis using ANVIL 4-0. (From Español & Shifres, 2015). The horizontal numerical scale (above) indicates the seconds. The first row shows the dynamic envelope of the sound signal. The following rows correspond to the variables of the Laban system (body, dimension, shape, effort, and flow). (Reproduced from Español & Shifres, 2015)

complementary rising form and a dabbing (direct-sudden-light) effort. Dabbing differs from punching in terms of weight; the conveyed feeling is then stimulated. S repeats this movement, emulating the double fall of the seat. Vocalizations also mimic the fall, and the soft timbre of her voice unites with the movement's qualities. First, the direct-sudden effort shawl movement joins the "Toc-toc" vocalization that keeps a similar duration. Second, very smooth (like a rocking) movements join the

second "Toc-toc" vocalization. The adult offers the baby a changed and elaborated repetition of the disturbing stimulus, interpreting it in a friendly nonverbal manner. The phrase closes with a freeze, silently and keeping eye contact to check that the baby has calmed down. The microanalysis of the sound and movement exhibits a hermeneutic intention in the adult's IDI-performance when confronting a disturbing stimulus. Expressive resources serve this goal, configuring a nonverbal hermeneutic.

2.3.2 Image-Schemas and Primary Metaphors

In this section, we reproduce a microanalysis of two moments extracted from a scene of an IDI-performance carried out by Martínez et al. (2018) to identify and characterize the allocation and features of the image-schematic forms and primary metaphors shaped and expressed in the IDI-performance. The selected scene corresponds to another moment of the adult–infant interaction extracted from the audiovisual register of the longitudinal case study mentioned in Sect. 3.1. The conditions, the participants Silvia (S) and Habib (H), and the infant's age (7 months old) are the same as in the previous analyses. The microanalysis of the scene was realized applying the Three-Component Multimodal Observational Model (Martínez, 2014) in which component (i) pursues the identification and description of the sequence of actions of the image-schematic form as performed, and components (ii) and (iii) aim at depicting the spatial, energetic, and temporal features of the multimodal modeling using similar tools as explained in the analyses above. Specific to our aims are the description of the adult–infant sound-linguistic expressions through which an image-schematic form and/or a primary metaphor is composed and enacted, and the type and quality (shape and effort) of the movements involved in the enaction of the image-schematic form and/or the primary metaphor enacted, respectively. For more details about the procedure and to access the whole microanalysis, see Martínez et al. (2018).

Brief Description of the General Scene

In this scene, S and H are seated on the floor facing each other. S produces, repeats, and varies sound and movement phrases while playing with H, and moves a pillow upward (pushing it away from H) and downward (approaching H). At the same time, S improvises verbal phrases and sings short cantilenas. The third repetition of the phrase ends with S approaching the pillow *as close to the baby's body as possible*. H smiles and shows facial gestures and movements denoting engagement. The interaction unfolds with permanent eye contact, only interrupted by the moving pillow. The phrase structure where the image-schematic forms are embedded is fully enacted with performative variations according to the repetition-variation form.

Moment 1 – Preparation and Enaction of the UP-DOWN Image-Schema

(i) *Preparation*

Brief Description
The initial moment of the scene consists of the adult's deployment of sound-kinetic actions to call the baby's attention. In pursuing this goal, S performs an UP-DOWN image-schematic form in the vertical axis (in this case in the DOWN-UP direction) consisting of pushing the pillow away from H to establish the initial location of what will be the beginning of the following image-schematic form.

Multimodal (Movement and Sound) Microanalysis of the Image-Schematic Form
S is seated in front of H holding a pillow at the DOWN location. She gently taps the baby's leg, moving her right hand in a rising-descending form with a dabbing effort (direct-sudden-light). S calls for H's attention saying: "Habibi…look!" using a whispered vocal emission in a high register. H – who was looking to one side – gently rotates his head and trunk, with a gliding effort (direct-sustained-light) looking at S's hand. Then, S performs a movement from DOWN to UP, raising the pillow with both arms until the object reaches a location above her head. This movement shapes a raising form with a gliding effort (direct-sustained-light). In sync with the pillow's ascendant movement, S says to H: "Look, look, look" with vocal emission in an ascendant pitch contour. Once the pillow reaches the UP location, she says: "Look where it is" using a whispering, high voice. The baby raises his head gently with a gliding effort (direct-sustained-light) directing to and holding his gaze to the pillow; he opens his mouth with a dubbing effort (direct-sudden-light) and smiles.

(ii) *First enaction of the SOURCE-PATH-GOAL image-schema*

Brief Description
The next moment in the scene consists of the adult's temporal expansion – through the sound-kinetic multimodal actions package – of the SOURCE-PATH-GOAL image-schematic form, anchored in the vertical axis of the Up-Down image-schema. In this way, both image-schemas are combined in S's IDI-performance. The three elements of this form are clearly marked with multimodal redundancy. After calling for H's attention at *SOURCE* – in the previous preparation moment – S enacts the *SOURCE-PATH-GOAL* image-schematic form producing vocal temporal and energetic variations in the vocal emission, and also varying the effort and shape of the movement. By moving the pillow toward the baby, and simultaneously saying: "And it comes to you," S arouses a sense of purpose in the infant at the *PATH* and a sense of fulfillment at the *GOAL* location.

Multimodal (Movement and Sound) Microanalysis of the Image-Schematic Form

*SOURCE*S
S is sitting in front of H. Holding the pillow with raised arms above her head at the UP location, she remains frozen for an instant and begins to sing a melodic motif

with the verbal phrase: "Acá arriba está, Habib"/"It is up here, Habib." The motif uses two pitches – separated by an interval of a minor third – that alternate rhythmically. Vocal emission is produced at a high register. The adult head's movement shapes the form advancing-retiring, with punching effort (direct-sudden-strong). The phrase concludes with a small movement, always at the top location – advancing the pillow with gliding effort (direct-sustained-light). While S calls for H's attention to the pillow at the *SOURCE* location, H gently raises his arms onto the pillow with floating effort (direct-sustained-light), smiling and looking to the pillow with expectation.

PATH

S moves the pillow from the UP to the DOWN location, approaching it to the baby and shaping a descending form with a gliding effort (direct-sustained-light). Simultaneously she pronounces the verbal phrase: "Y viene para Habib"/"It is coming towards Habibi" with a descendant vocal emission, encompassing an interval of an octave.

GOAL

Once the pillow arrives at the DOWN location, the adult approaches the pillow *as close as possible* to the baby's body. H grasps the pillow with an enclosing form with gliding effort (direct-sustained-light) and leans on it, remaining in contact with the pillow.

Moment 2 – CODA, with Activation of the Primary Metaphor INTIMACY IS PROXIMITY

Brief Description

After the third repetition of the SOURCE-PATH-GOAL image-schematic form (performed with expressive variations), the last moment (called CODA) is added. It reaches the climax of the whole IDI-performance. When the climax is achieved at the GOAL location, and the adult approaches her head to the baby's – making contact with the pillow – the primary metaphor INTIMACY IS PROXIMITY is activated. At that moment, both domains (proximity and intimacy) are co-activated, merging the experience of intimacy with that of feeling physically close with affection.

Multimodal (Movement and Sound) Microanalysis

Seated in front of the baby at the DOWN location, the adult softly begins to withdraw the pillow, moving the object in the horizontal plane with gliding effort (direct-sustained-light). While releasing the pillow from the baby's body, S sings: "Y viene para Habib"/"It is coming toward Habibi" *pianissimo* and *rallentando*. S moves the pillow switching sides – always at DOWN position at the floor – flexing her trunk toward H, approaching her head to the baby's, and laughing; the scrubbing-on-the-pillow movement is performed in the horizontal axis with dubbing effort (direct-sudden-light). Then, she withdraws, and her movement shapes a composed advancing-retiring and descending-rising form, with a gliding effort

(direct-sustained-light). H moves his head again softly, in smooth contact with the pillow. In the end, S lines again approaching her head to the baby's and articulates: "Uh, uh, je je je je" in ascendant-descendant contour.

After the climax is achieved at the GOAL location and the CODA takes place, and the adult approaches her head to the baby's making contact with the pillow, the primary metaphor INTIMACY IS PROXIMITY (related to the primary experience of being physically close to the people with whom we have an intimate relationship) is activated. Both domains (proximity and intimacy) are co-activated, merging the experience of intimacy with that of feeling physically close with affection. The conflation of the three components intimacy-closeness-contact displays an appropriate affective context for the activation of this primary metaphor experience.

Therefore, image-schematic forms and primary metaphors are embedded in the multimodal packages that the adult offers to the baby; multimodal redundancy contributes to highlighting elements of the image-schematic forms enacted by the adult to and for the baby, and the affective content of experience is aesthetically shaped throughout the elaboration of the vitality forms contour in the activation of primary metaphors (i.e., *the ways* the adult moves the pillow toward and forward the baby until arriving at the climax of closeness and intimacy, the effort when the heads touch each other, the level of energy of the adult's laugh at the climax, and so on).

2.4 What Happens Through the IDI-Performances?

IDI-performances are sound-kinetic phrases improvised through the repetition of simple motifs. Motifs are repeated in a varied way when they are molded spatially, energetically, and temporally. IDI-performances emerge through the resources also used in temporal art performances: the repetition-variation form and variation modeling. Although these two resources are not exclusive to IDI-performances, they organize them and converge in them.

(i) *The adult summons the infant to social life and supports him through structural and expressive resources also used in the performing arts***.** The varied repetition of simple motifs is a powerful resource to prolong attention, support, and strengthen intersubjective engagement. Studies on IDS indicate that the use of the sound repetition-variation structure reinforces the infant's attention. Our analyses go beyond, showing that in IDI-performances movement is the heart of the repetition-variation form. In Sect. 2.3.1. Moment 1, repeating a simple kinetic motif like removing the blanket, fills the infant with expectation and promotes anticipation so he can guess what comes next. The slightly different varied repetition leads to complex and striking multimodal modeling and novelty, like when the gliding movement of the shawl's removal combines with sharp, regular, short sounds (staccatos) "Ti-ti-ti-ti…." The adult *invents* the motif and attunes with the baby to playfully model it. With each repetition, she embellishes or adds attractiveness using expressive

resources, provided the infant is willing to receive and respond. The timing seems to be the other crucial aspect of the IDI-performance: sound and movement duration (shortening and lengthening) are major resources to increase expectation and sustain attention. The IDI-performance lengthens the sharing moment, letting the infant experience an array of forms of vitality he would not be able to access by himself. Through IDI-performances, adults and infants enjoy together and share the vitality forms. (For further details on forms of vitality, see Chap. 9.)

(ii) *There is an expressive adult improvisation that turns each exchange into a one-of-a-kind IDI-performance.* When gathering with the infant, the adult uses handy personal and cultural skills to create IDI-performances congruent with the infant's response and the contingencies of the environment. This requires that the use of such resources be flexible and dynamic. In this sense, the adult behaves like an improviser capable of configuring expressive statements where form and content are always novel. There are traditions of musical and theatrical performance where improvisers focus on the expressions of their audience to engage in their actions and ensure long-lasting performances (Zeranska-Kominek, 1992). In the same way, adult improvisation is fundamental to guarantee the exchange: his creation unfolds in the double aspect of his composition and the infant's reception of it. An expressive mode of operation on the temporal, energetic, and spatial patterns that organize and color adult-to-infant directed behavior makes each exchange a one-of-a-kind IDI-performance. On the other hand, IDI-performance improvisation eases the articulation with the outside world: the irruption of a loud sound and a rough movement (such as the chair's fall) can be incorporated into the IDI-performance (instead of causing the interruption) and become part of the everyday environment. Improvisation thus allows the organization of the repetition-variation form to be flexible and adaptable to contingency. Improvisation is prototypical in ordinary IDI-performances and gives shape to those set in ritualized culture-dependent structures (such as the peek-a-boo in Sect. 2.3.1. Moment 2).

(iii) *The adult offers well-formed, conspicuous behavioral units with clear start and end markers*. Adults deliver holistic well-performed sound-movement attractive units. The reiteration of simple motifs brings unity to the IDI-performance. In turn, adults outline the boundaries of that coherence. They initiate the IDI-performance upon perceiving they have the infants' full attention. In Sect. 3.1. Moment 1, S makes small movements to catch the baby's eye, but when this turns unsuccessful, she slowly starts covering and uncovering him with the shawl waiting for a response. When the baby turns his head, the IDI-performance develops. The adult also usually marks the closure. By prolonging movements, resorting to *rubatos*, returning to the center of her kinesphere (as in Sect. 2.3.1, Moment 1), or by remaining still and silent and upholding eye contact (as in Sect. 2.3.1, Moment 2). Adults deliver holistic attractive units organized through a motif they selected from an infinite repertoire, usually reciprocal with some behavior, state, or event close to or in

intimate contact with the infant, and motif repetitions embellished through structural and expressive resources. These attractive, well-performed, and well-delimited phrases that emerge in mutuality contexts may induce the recognition and processing of meaningful units within the information flow continuum, favoring their detection, which is a crucial aspect of cognitive development.

(iv) *The adult's organization of sounds and movements promotes the simultaneous experience of separation/continuity in time***.** The repetition-variation structure organizes the passage of time. Structuring through the configuration of sound and movement phrases does not compartmentalize time or offer "time bricks"; on the contrary, a mutual articulation of sound and movement guarantees fluidity and the passage from one meaning unit to another. The expressive resources serve this demarcation/continuity. For example, using ritardando in the vocal emission of phrase endings demarcates an end, while new movements project toward what comes next. Thus, in Sect. 2.3.1, the vocalizations form an *arc* in terms of melodic contour, initially ascending and then descending. While the descending melodic arc ends the phrase, the adult starts the movement with the blanket, anticipating the peek-a-boo play. Similarly, in the section we call *A fortuitous event* (see Sect. 2.3.1), the IDI-performance organizes around an underlying pulse and the vocalizations adapt. When this moment ends, she takes another vocal motif (used later to build the next IDI-performance): "*Esa* ca..." (an apocopate of "*Esa carita"*) that becomes the main idea in the following sentence. The motif *"Esa* ca..." is thematically part of the subsequent phrase, but structurally speaking, it belongs to the previous one. The adult seems to think of the next section's motif while still finalizing the structure (rhythmic and melodic) of the previous one. The baby is thus driven in time, sometimes by sound, and sometimes by movement. But frequently, both modalities complement each other, particularly in the demarcation of large units (e.g., in Sect. 2.3.1, Moment 1, the third movement repetition extends with a sound and movement ritardando).

(v) *The adult softens or interprets the world for the infant and regulates his emotional states***.** The analysis of sound and movement uncovers some hermeneutic intention in adult IDI-performance when facing disruptive stimuli. This happens through the IDI-performances and other multimodal packages. In Sect. 2.3.1, Moment 1, the stressed pronunciation of the word "teléfono" (telephone) conforming to a pulse (a multimodal package without the repetition-variation form) reduces the tension, provides a safety message, and helps regain a serene atmosphere. In Sect. 2.3.1, Moment 2, when the seat suddenly fell, the adult improvised an IDI-performance softly emulating the fall, offering a modified and elaborated repetition of the disruptive stimulus. In the first case, adjusting the events to a regular pulse helps regain contact when jeopardized by the impact of external stimuli. On the contrary, when the intersubjective interchange occurs without external interference, and the infant focuses on the adult's IDI-performance, the pulse becomes more

flexible. In the second case, the IDI-performance offers a nonverbal interpretation conveying the message that the outside stimuli are now friendlier.

(vi) *The adult illuminates image-schemas.* Participating in IDI-performances clears the way to the gestaltic experiences of image-schemas in inter-enactive contexts that will nurture posterior individual sensory-motor infant experiences. Image-schemas are imbued and highlighted in the artistic design of IDI-performances: the components of the image-schematic form are underlined through the vitality forms that adults display for-and-with the infants using expressive resources and multimodal redundancy. As we noticed in the analysis of Moment 1 (Sect. 2.3.2), the three components of the SOURCE-PATH-GOAL image-schema are multimodally stressed in S's IDI-performance: i.e., she produces successive upward-backward-expansion-contraction movements emphasizing the SOURCE location. The infant's embodied experience – linked to image-schemas – shapes specific features in early social cognition. These embodied traits take form when the image-schema unfolds in the moving adult and the infant is observing, or when he can sense it proprioceptively while being held in the arms of another person.

(vii) *The adult invites the infant to intersubjectively-intercorporeally experience primary metaphors with her***.** Primary metaphors are embedded in the structure of the image-schematic forms in IDI-performance. The affective dimension is activated during the image-schematic progression, conflating cognition and affection. As we showed in Moment 1 (Sect. 2.3.2), the primary metaphor CLOSENESS IS INTIMACY – activated at the climax when the adult approaches her head to the baby's making contact with the pillow – unveils H's primary experience of being physically close to S, enduring an intimate relationship. The affective situation also highlights the rewarding feeling of reaching the GOAL. Ultimately, this intimate encounter counts as a trace of affective sociality.

(viii) *Through IDI-performances, the adult shows a ludic attitude.* Almost from the beginning of the baby's life, adults approach them playfully. At around the second month, early social play emerges in the dyad. This play is joyful in nature and has social exchange and enjoyment as its sole purpose (Stern, 1985, 2010). Early social play brims with musicality and involves the loving experience of being together or mutuality. IDI-performances are an early mode of social play, like imitation cycles. Unlike other types of early play, IDI-performances are an intrinsically playful way of convening social life; non-playful versions of IDI-performances are not possible, but there can be non-playful versions of imitations or protoconversations. The IDI-performances serve no other purpose than playful togetherness (even when they repair or soften a situation). The ludic attitude of an adult initiating early social play will accompany the child throughout play development, scaffolding different infant capacities. But what is special in IDI-performances is that adults expose vitality forms, play with them, inviting the infant to participate

in a key aspect of the interpersonal world and the arts from the very beginning of life (see Chap. 9 for continuity on this subject).

(ix) *Through the IDI-performances the baby's aesthetic and body enculturation begins.* In our cultural context of complex stratified societies, artistic activities appear primarily as a spectacle, that is, as something to be contemplated that can attract attention and affect our emotions by infusing delight, astonishment, pain, or other affections. In this way, the experience of early interactions could be connected with those of artistic reception in adult life, particularly music and dance as entertainment, typical of our societies, as suggested by Stern (1985). We might guess that in societies where the arts are not culturally conceived as a spectacle, early interactions might be experienced differently. Besides, as we previously mentioned, some expressive cultural features of sound and movement are incorporated into the repertoire of resources that adults use in reciprocity experiences with their infants. For example, a pulse the adult adjusts to attract the infant's attention and regulate uncertainty can be found in the IDI-performance. However, it is also clear that the adult diverts from this pulse to direct the focus toward a particular speech point, to intensify the enjoyment and the state of mutuality, and to sustain attention for a longer period. This use of timing reflects the functionality of the resource in artistic practices of culture. We can speculate that cultural features are unconsciously conveyed through the use of such expressive resources. The detection of expressive resources proper to each culture in IDI-performances urges us to recognize that intuitive parenting comprises cultural contents. The way of protecting, nurturing, and stimulating are traversed by cultural contents and are a primary instance of education of cultural aspects.

It is a clear example of an assembly of general resources of the species coupled with behaviors that characterize its cultural framework. Finally, through IDI-performances, infants are exposed to the bodily modes of their culture. There are culture-specific movement styles (with variations in its rhythmicity, in the amplitude and speed of movements, in its sound/kinetic contours). These modes pervade the entire care process, from how the baby is held, to how he is bathed and cuddled. And they are highlighted in the IDI-performances that embellish, exaggerate, and reiterate them in a context where no other activity is foregrounded: it is only about sharing and enjoying together. This repeated and playful exposure to resources shared with music and dance (fully embodied arts) is one way by which infants begin to connect with the embodied arts of their culture and their corporeality begins to be shaped. Very little is known about this last issue. Some clues on how culturally prescribed moving and expressing patterns shape the baby's corporeality can be found in Chap. 5.

2.5 Final Remarks

The aesthetic perspective of dyadic interaction assumed in this work led us to continue the path initiated by others in evidencing remarkable structural and functional affinities between intersubjective encounters in infancy and communicational-affective experiences in the temporal arts. It allowed us to light up the adult unconscious, unintentional capacity for improvisation that emerges when an adult meets a baby. An ability to improvise sustained by a simple strategy at the serve of reciprocity: the repetition-variation form. By using analytical tools and concepts developed for the exegesis of artistic expressions, we arrived at a clear-cut description of IDI-performance. These analytical tools also allowed us to show the initial joint experiential matrix of affect and cognition and the other many fascinating events that occur through an IDI-performance.

Our work provides a route for exploration and therefore it still has some limitations. There is much yet to be done. We have focused on interactions with 7-month-old infants. It is necessary to investigate whether structural aspects detected in IDI-performances and their functional value vary or remain stable throughout the first year of life. It is feasible and relevant to keep exploring other structural and functional aspects of adult–infant interaction compatible with those of music and dance. Microanalysis provides information that naturalistic observation in real-time cannot give. The wealth of information provided by microanalysis offsets the problem of the generalization of the results, but certainly, the study of the subject should be supplemented by other methodological approaches. We have observed and commented on infant behavior, but our work focused mostly on adult IDI-performance. The hypothesis about the functional value of the IDI-performance requires further research addressing infant participation. Our study concentrated on this particular mode of adult–infant interaction; it would be interesting to see how it relates to other dyadic exchanges. The participants of the microanalytic studies discussed belonged to our culture. It would be worth exploring dyads from other cultures that allow us to ponder the scope of this form of interaction and the particular features it assumes in different cultural contexts. In addition, it would also be valuable to know their trajectory, and whether they are abandoned or remain at play in the course of development.

This framework developed in this chapter on IDI-performance serves as the theoretical background for further research. Much of what remains to be done is addressed in the following chapters. The remainder of the book contains ideas of IDI-performances development, about their relationship with protoconversations and action songs during the first year of life (see Chap. 3), and their link with *forms of vitality play*, a non-figurative play that emerge during the third year of life (see Chap. 9). Additionally, in Chap. 6, we use the description of the IDI-performance to design experimental stimuli that mimic parental IDI-performance.

The value and the poietic effect of the corporeal and aesthetic dimension of human life light up in IDI-performances. Body and art come together. Perhaps this union explains why being with babies is a good thing for adults. To be with babies

is to participate in an intersubjective experience that takes the form of a body-to-body encounter – along with the abundance of sensations and sensitivities it entails – full of musicality, reciprocity, and kinetic sensitivity.

References

Austin, J. L. (1962/1975). In M. Sbisà & J. O. Urmson (Eds.), *How to do things with words* (2nd ed.). Oxford University Press.

Bahrick, L., & Hollich, G. (2008). Intermodal perception. In M. Haith & J. Benson (Eds.), *Encyclopedia of infant and early childhood development* (2nd ed., pp. 164–176). Academic Press.

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. Academic Press Inc.

Bateson, M. C. (1979). The epigenesis of conversational interaction: A personal account of research development. In M. Bullowa (Ed.), *Before speech. The beginning of interpersonal communication* (pp. 63–78). Cambridge University Press.

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother–infant face-to-face communication. *Developmental Psychology, 52*(4), 556. https://doi.org/10.1037/a0040085

Brand, R. J., Baldwin, D. A., & Ashburn, L. A. (2002). Evidence for 'motionese': Modifications in mothers' infant-directed action. *Developmental Science, 5*, 72–83. https://doi.org/10.1111/1467-7687.00211

Bråten, S. (1998). Infant learning by altercentric participation: The reverse of egocentric observation in autism. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 105–124). Cambridge University Press.

Calvo-Merino, B. (2010). Neural mechanisms for seeing dance. In B. Bläsing, M. Puttke, & T. Schack (Eds.), *The neurocognition of dance: Mind, movement and motor skills* (pp. 153–176). Psychology Press.

Condon, W. S., & Sanders, L. W. (1974). Neonate movement is synchronized with adult speech: Interactional participation and language acquisition. *Science, 183*(4120), 99–101. https://doi.org/10.1126/science.183.4120.99

Davies, E. (2006). *Beyond dance. Laban's legacy of movement analysis*. Routledge.

Dissanayake, E. (2000). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music* (pp. 389–410). The MIT Press.

Dissanayake, E. (2001). Becoming homo Aestheticus: Sources of aesthetic imagination in mother-infant interactions. *SubStance, 30*(1/2), 85–103. https://doi.org/10.2307/3685506

Dissanayake, E. (2008). The arts after Darwin: Does art have an origin and adaptive function? In K. Zijlmans & W. van Damme (Eds.), *World art studies: Exploring concepts and approaches* (pp. 241–263). Valiz.

Dissanayake, E. (2017). Roots and route of the artification hypothesis. *AVANT. Pismo Awangardy Filozoficzno-Naukowej, 1*, 15–32.. https://doi.org/10.26913/80102017.0101.0001.

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana [How to do things without words. Gesture and fiction in early infancy]*. Antonio Machado.

Español, S. (2010). Performances en la infancia: cuando el habla parece música, danza y poesía [Performances in infancy: when speech seems music, dance and poetry]. *Epistemus. Revista de Estudios en Música, Cognición y Cultura, 1*(1), 57–95. https://doi.org/10.21932/epistemus.1.2702.0

Español, S. (2014). La forma repetición-variación: una estrategia para la reciprocidad [The repetition-variation form: a strategy for reciprocity]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana [Music and developmental psychology. An interdisciplinary exploration of human musicality]* (pp. 157–192). Paidós.

Español, S., & Shifres, F. (2015). The artistic infant directed performances: A mycroanalysis of the adult's movements and sounds. *Integrative Psychological and Behavioral Science, 49*(3), 371–397. https://doi.org/10.1007/s12124-015-9308-4

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60*, 1497–5100. https://doi.org/10.2307/1130938

Ferrari, P. F., & Gallese, V. (2007). Mirror neurons and intersubjectivity. In S. Bråten (Ed.), *On being moved: From mirror neurons to empathy* (pp. 73–88). John Benjamins Publishing.

Freedberg, D., & Gallese, V. (2007). Motion, emotion and empathy in esthetic experience. *Trends in Cognitive Sciences, 11*(5), 197–203. https://doi.org/10.1016/j.tics.2007.02.003

Gabrielsson, A. (2003). Music performances research at the millennium. *Psychology of Music, 31*(3), 221–272. https://doi.org/10.1177/03057356030313002

Gabrielsson, A., & Lindström, E. (2010). The role of structure in the musical expression of emotions. In P. Juslin & J. A. Soloboda (Eds.), *Handbook of music and emotion: Theory, research, applications* (pp. 367–400). Oxford University Press.

Gallese, V. (2017). Visions of the body. Embodied simulation and aesthetic experience. *Aisthesis, 10*(1), 41–50. https://doi.org/10.13128/Aisthesis-20902

Gratier, M., & Apter-Danon, G. (2009). The improvised musicality of belonging: Repetition and variations in mother-infant vocal interaction. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 281–300). Oxford University Press.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology, 6*(1167), 236–245.

Hoch, J. E., Ossmy, O., Cole, W. G., Hasan, S., & Adolph, K. E. (2021). 'Dancing' together: Infant–mother locomotor synchrony. *Child Development, 92*(4), 1337–1353. https://doi.org/10.1111/cdev.13513

Imberty, M. (2002). La musica e il bambino. In J.-J. Nattiez (Ed.), *Enciclopedia della musica* (pp. 477–495). Giulio Einaudi Editore.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development, 66*, i–149.

Johnson, M. (1987). *The body in the mind*. University of Chicago Press.

Johnson, M. (2007). *The meaning of the body*. University of Chicago Press.

Jones, A. (2012). Foreword: Kinesthetic empathy in philosophical and art history: Thought on how and what art means. In D. Reynolds & M. Reason (Eds.), *Kinesthetic empathy in creative and cultural practices*. Intellect.

Juslin, P. N. (2013). What does music express? Basic emotions and beyond. *Frontiers in Psychology, 4*, 596. https://doi.org/10.3389/fpsyg.2013.00596

Kalashnikova, M., Carignan, C., & Burnham, D. (2017). The origins of babytalk: Smiling, teaching or social convergence? *Royal Society Open Science, 4*(8), 170306. https://doi.org/10.1098/rsos.170306

Kaye, K., & Fogel, A. (1980). The temporal structure of face-to-face communication between mothers and infants. *Developmental Psychology, 16*(5), 454. https://doi.org/10.1037/0012-1649.16.5.454

Kendall, R. A., & Carterette, E. C. (1990). The communication of musical expression. *Music Perception, 8*(2), 129–163. https://doi.org/10.2307/40285493

Kipp, M. (2004). *Anvil – A video annotation research tool*. http://www.dfki.de/~kipp/anvil. June 2008.

Koterba, E. A., & Iverson, J. M. (2009). Investigating motionese: The effect of infant-directed action on infants' attention and object exploration. *Infant Behavior and Development, 32*(4), 437–444. https://doi.org/10.1016/j.infbeh.2009.07.003

Laban, R. (1971). *The mastery of movement*. Plays.

Lakoff, G. (2008). The neural theory of metaphor. In R. Gibbs (Ed.), *The Cambridge handbook of metaphor and thought* (pp. 17–38). Cambridge University Press.

Lakoff, G., & Johnson, M. (1980/2003). *Metaphors we live by*. University of Chicago Press.

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh*. Basic Books.

Lansdale, J. (2009). Preface to Moore, Carol-Lynne. In *The harmonic structure of movement, music, and dance according to Rudolf Laban: An examination of his unpublished writings and drawings* (pp. xi–xiii). Edwin Mellen Press.

Levinson, S. C., & Torreira, F. (2015). Timing in turn-taking and its implications for processing models of language. *Frontiers in Psychology, 6*, 731. https://doi.org/10.3389/fpsyg.2015.00731

Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4-to 10-month-old infants. *Developmental Psychobiology, 48*(4), 288–300. https://doi.org/10.1002/dev.20140

Licata, M., Paulus, M., Thoermer, C., Kristen, S., Woodward, A., & Sodian, B. (2014). Mother–infant interaction quality and infants' ability to encode actions as goal-directed. *Social Development, 23*(2), 340–356. https://doi.org/10.1111/sode.12057

Malloch, S. (1999/2000). Mothers and infants and communicative musicality. *Musicæ Scientiæ*, Special Issue, 29–57. https://doi.org/10.1177/10298649000030S104

Malloch, S., & Trevarthen, C. (2009). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 1–10). Oxford University.

Mandler, J. M. (2005). How to build a baby III: Image-schemas and the transition to verbal thought. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 137–163). Mouton de Gruyter.

Markova, G., & Legerstee, M. (2006). Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology, 42*(1), 132. https://psycnet.apa.org/doi/10.1037/0012-1649.42.1.132

Martínez, I. (2014). La base corporeizada del significado musical [The embodied basis of musical meaning]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana [Music and developmental psychology. An interdisciplinary exploration of human musicality]* (pp. 71–110). Paidós.

Martínez, I., Español, S., & Pérez, D. (2018). The interactive origin and the aesthetic modelling of image-schemas and primary metaphors. *Integrative Psychological and Behavioral Science, 52*(4), 646–671. https://doi.org/10.1007/s12124-018-9432-z

Miall, D., & Dissanayake, E. (2003). The poetics of Babytalk. *Human Nature, 14*(4), 337–364. https://doi.org/10.1007/s12110-003-1010-4

Moore, C. L. (2009). *The harmonic structure of movement, music, and dance according to Rudolf Laban: An examination of his unpublished writings and drawings*. Edwin Mellen Press.

Newlove, J. (2007). *Laban for actors and dancers. Putting Laban's movement theory into practice: A step-by- step guide*. Routledge.

Nudler, A. C., Jacquier, M. d. l. P., & Español, S. (2020). Las formas de la vitalidad en la recepción de una obra teatral [Forms of vitality in the reception of a theater work]. *Epistemus. Revista de Estudios en Música, Cognición y Cultura, 8*(1), 2–24. https://doi.org/10.24215/18530494e013

Papoušek, H. (1996a). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliege & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 37–55). Oxford University Press.

Papoušek, M. (1996b). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 88–112). Oxford University Press.

Papoušek, M., & Papoušek, H. (2002). Intuitive parenting. In M. H. Bornstein (Ed.), *Handbook of parenting. Volume 2. Biology and ecology of parenting* (pp. 183–203). Lawrence Erlbaum.

Provenzi, L., Scotto di Minico, G., Giusti, L., Guida, E., & Müller, M. (2018). Disentangling the dyadic dance: Theoretical, methodological and outcomes systematic review of mother-infant dyadic processes. *Frontiers in Psychology, 9*, 348. https://doi.org/10.3389/fpsyg.2018.00348

Reddy, V. (2005). Before the 'third element': Understanding attention to self. In N. Eilan, C. Hoerl, T. McCormack, & J. Roessler (Eds.), *Joint attention: Communication and other minds* (pp. 85–109). Oxford University Press.

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Reddy, V., Chisholm, V., Forrester, D., Conforti, M., & Maniatopoulou, D. (2007). Facing the perfect contingency: Interactions with the self at 2 and 3 months. *Infant Behavior and Development, 30*(2), 195–212. https://doi.org/10.1016/j.infbeh.2007.02.009

Reynolds, D. (2012). Kinesthetic empathy and the dance's body: From emotion to affect. In D. Reynolds & M. Reason (Eds.), *Kinesthetic empathy in creative and cultural practices* (pp. 121–136). Intellect.

Rivière, A. (1986/2003). Interacción precoz. Una perspectiva vygotskiana a partir de los esquemas de Piaget [Early interaction. A Vygotskian perspective from Piaget's schemas]. In M. Belinchón, A. Rosa, M. Sotillo, & I. Marichalar (Eds.), *Ángel Rivière. Obras Escogidas [Ángel Rivière. Selected works]* (Vol. Vol II, pp. 109–142). Panamericana.

Sheets-Johnstone, M. (2011). *The primacy of movement* (Vol. 82). John Benjamins Publishing.

Sheets-Johnstone, M. (2015). *The corporeal turn: An interdisciplinary reader*. Andrews UK Limited.

Shifres, F. (2014). Algo más sobre el enlace entre la infancia temprana y la música. El poder expresivo del *rubato* [Something more about the link between early childhood and music. The expressive power of rubato]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana [Music and developmental psychology. An interdisciplinary exploration of human musicality]* (pp. 21–70). Paidós.

Shifres, F. (2020). El tiempo como experiencia compartida [Time as a shared experience]. In C. Fonseca (Ed.), *Ritmicidades. Cuerpos en JIRA (2015–2019) [Rhythmicities. Bodies in ICRA (2015–2019)]* (pp. 82–97). Universidad Nacional de las Artes.

Slater, A., Quinn, P. C., Kelly, D. J., Lee, K., Longmore, C. A., McDonald, P. R., & Pascalis, O. (2010). The shaping of the face space in early infancy: Becoming a native face processor. *Child Development Perspectives, 4*(3), 205–211. https://doi.org/10.1111/j.1750-8606.2010.00147.x

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187–213). Wiley-Interscience.

Stern, D. N. (1985). *The interpersonal world of the infant. A view from psychoanalysis and developmental psychology*. Basic Books.

Stern, D. N. (2010). *Forms of vitality. Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford University Press.

Stern, D. N., Beebe, B., Jaffe, J., & Bennet, S. L. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 177–202). Academic.

Trehub, S. E. (2010). In the beginning: A brief history of infant music perception. *Musicae Scientiae, 14*(2_suppl), 71–87. https://doi.org/10.1177/2F10298649100140S206

Trehub, S. E. (2015). Infant musicality. In *The Oxford handbook of music psychology* (pp. 387–398). Oxford University Press.

Trevarthen, C. (1982). The primary motives for cooperative understanding. In G. Butterworth & P. Light (Eds.), *Social cognition* (pp. 77–109). Harverster.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). Cambridge University Press.

Trevarthen, C. (2017). The function of emotions in early infant communication and development. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 48–81). Routledge.

Trevarthen, C., & Reddy, V. (2007). Consciousness in infants. In M. Velmans & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 41–57). Blackwell Publishing.

Tronick, E., Als, H., Adamson, L., Wise, S., & Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *Journal of the American Academy of Child Psychiatry, 17*(1), 1–13. https://doi.org/10.1016/s0002-7138(09)62273-1

Weinberg, M. K., & Tronick, E. Z. (1996). Infant affective reactions to the resumption of maternal interaction after the still-face. *Child Development, 67*(3), 905–914. https://doi.org/10.2307/1131869

Zeranska-Kominek, S. (1992). Mode: Process and/or structure? A study of the Tukmen mukam Gökdepe. In R. Dalmonte & M. Baroni (Eds.), *Secondo convegno europeo di analisi musicali [Second European conference on music analysis]* (pp. 249–258). Universitá degli Studi di Trento.

Silvia Español holds a PhD in Psychology from Universidad Autónoma de Madrid (Spain). She is currently a researcher at the Latin America Institute of Social Research (IICSAL), Faculty of Latin America Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET, Argentina). She coordinates the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc in Cognitive Psychology, and Director of FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is a certified Somatic Educator in the Feldenkrais Method. Her work deals with cognitive developmental psychology, psychology of music, and the study of human movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/silvia.espanol

Favio Shifres is a pianist, chamber music teacher and orchestral conductor from the National University of La Plata (UNLP, Argentina), and PhD from the University of Roehampton (United Kingdom). He is a regular Professor in Auditory Education and Comparative Music Education (UNLP), and Psychology of Music and Arts in several postgraduate courses (Universidad nacional dde las Artes, UNA; FLACSO; UNLP, Universidad Nacional de Rosario, UNR). Director of research projects. Editor in Chief of Epistemus. Journal of music, cognition and culture. Guest editor in numerous research journals. Founding member and former president of the Argentine Society for the Cognitive Sciences of Music (SACCoM). He has written numerous papers in her area of expertise: https://www.aacademica.org/favio.shifres

Isabel Martínez has a PhD in Psychology of Music from University of Surrey, United Kingdom. She is Professor of Music (Music Department, Faculty of Arts) and researcher (Laboratory for the Study of Musical Experience LEEM-FBA) at the Universidad Nacional de La Plata (UNLP, Argentina). She is First President of the Argentine Society for the Cognitive Sciences of Music (SACCoM). She is currently Director of the Laboratory for the Study of Musical Experience (LEEM-FBA-UNLP). She develops research in the field of embodied music cognition in topics that links psychology of music, cognitive developmental psychology, and philosophy of mind. In 2017 she received the UNLP Award for Scientific, Technological and Artistic Work. She has written numerous papers in her area of expertise: https://www.researchgate.net/profile/Isabel-Martinez-5

Diana Pérez has a PhD in Philosophy at the Universidad de Buenos Aires. She is the Director of the Philosophical Research Institute (IIF-SADAF-CONICET); Full Professor of Metaphysics at the Philosophy Department, and Faculty member of the MA in Cognitive Psychology, Universidad de Buenos Aires. Her main areas of research are the metaphysics of the mind, philosophy of cognitive sciences and philosophy of art. She published three books: La mente como eslabón causal [Mind as a causal link] (1999), Sentir, desear, creer [Feeling, desiring, believing. A philosophical Approach to psychological concepts] (2013), Social Cognition and the Second Person in Human Interaction (with A. Gomila) (2021), and also edited other seven books as long as numerous papers on philosophy of mind, metaphysics, and philosophy of art. She gave talks and participate in conferences in Germany, Brazil, Chile, Colombia, USA, France, Spain, Mexico, Peru and Uruguay. (https://iif.conicet.gov.ar/researchers-en/researchers-conicet-en/diana-perez-en/).

Chapter 3
Infant-Directed Improvised Performances, Protoconversations, and Action Songs During the First Year of Life

Soledad Carretero, Silvia Español, Fernando G. Rodríguez, and Favio Shifres

Abstract The connection between babies and adults is permeated by a communicative musicality that manifests in different ways. Three of these manifestations are relevant to social cognitive development: infant-directed improvised performances, protoconversations, and action songs. They typify the first social and multimodal exchanges that allow adults and babies to share moments of deep intimacy and affection. While protoconversations are extensively studied interaction in developmental psychology and action songs are widely known, infant-directed improvised performances have only recently been identified. The three variants of these exchanges are brought together as a unitary category under the name of emergent organizations of early communicative musicality. The chapter presents a longitudinal case study of a mother–baby dyad between the 2nd and 9th month of the baby, analyzing the structure, frequency, and developmental trajectory of these three emergent organizations. It also describes the various ways in which multimodal stimulation is organized in each of them.

S. Carretero
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina, Buenos Aires, Argentina

S. Español (✉)
Latin American Institute of Social Research (IICSAL), Latin American Faculty of Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina
e-mail: sespanol@flacso.org.ar

F. G. Rodríguez
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina, Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

F. Shifres
Laboratory for the Study of Musical Experience, Faculty of Fine Arts, Universidad Nacional de La Plata (LEEM-FBA-UNLP), La Plata, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_3

Keywords Protoconversations · Action song · Communicative musicality · Infant-directed improvised performance · Dance · Music

3.1 Introduction

Within the framework of embodiment, some subfields of cognitive sciences have undergone a paradigmatic revolution by conceiving cognitive processes considering the emotional, perceptual, cognitive and motor experiences of the subject in the environment (Calvo & Gomila, 2008, Johnson, 2007, Leman, 2008; Durt, 2017). Concomitantly, the corporeal turn (Sheets-Johnstone, 2011, 2015) advocates the reconceptualization of psychology and, also of philosophy, placing corporality in movement at its center. The long-overlooked relevance of the body for the understanding of human processes is nowadays vindicated in the most comprehensive terms.

Within this framework, interactions between adults and infants have been studied by relocating the focus on different modes of bodily contact and their progressive semiotization. Psychologists, on the fringes of the cognitivist frenzy and concentrating on the earliest stages of life (where they found the use of adult individual categories unhelpful), showed clear interest toward what happened in the corporeally mediated exchange between an adult and an infant or young child (Trevarthen, 1998; Stern, 1985, 2010; Reddy, 2008; Fogel et al., 1988; Fogel & Garvey, 2007; Garvey & Fogel, 2007). It was evident that encounter and mutual understanding during these exchanges depended on interactions involving characteristics that differ from those that organize our everyday adult lives. Given their priority in development, these characteristics shape the fundamental interhuman platform. Work on infant healthy development specifically emphasizes this original contact as a decisive factor (Hernandez-Reif et al., 2007).

The first human relationship involves two unequal protagonists, adult and infant. Language divides the participants who establish a behavioral and emotional back and forth. Fortunately, babies are born in the arms of a conspecific whose empathy and affective attunement abilities make him or her sensitive to their needs. These abilities underlie what has been called intuitive parenting: a spontaneous and overall capacity to protect, feed, stimulate, and teach infants the basics of their cultural world (Papoušek, 1996a; Papoušek & Papoušek, 2002). In fact, adults act differently with infants than they do with other adults. For example, when speaking to their infants, they use increased tonal variations, hyperarticulation of vowels, repetitions and short phrases, exaggeration of prosodic features, prototypical melodic contours, etc. (Fernald, 1989; Papoušek, 1996b). These alterations of speech and musicality that characterize the way adults act toward infants are potential conditions for the establishment of early intersubjective experiences (Trevarthen, 2000; Imberty, 2002; Trevarthen & Reddy, 2007).

In these interactions, a corporeal and temporal organized way has been acknowledged through which babies and adults are able to attune their interaction and

affective states: they connect through a communicative musicality, coupling in their sonorous and motor behaviors. The constitutive dimensions of this musicality are pulse, quality, and narrative. While interacting with others, these three components foster encounter and coordination. The pulse determines the way actions are temporally patterned by the establishment of regularities. In contrast, quality refers to the non-temporal attributes of sound and movement: shape and contours for the latter, spectrum and pitch for the former. The narrative concerns the organization of action sequences acquiring a sense of unity through their form and content (for a detailed presentation, see Chap. 2). Research on communicative musicality encompasses the study of adult behavior in ritual events, dance and musical performance, and the study of temporal arts – music and dance – or antecedents of temporal art in infancy and early childhood (Gratier & Apter-Danon, 2009, Gratier & Trevarthen, 2008, Malloch & Trevarthen, 2009, see also Chap. 2 in this volume). Trevarthen and Reddy (2007) have hypothesized that this type of interaction starts unfolding by the age of 2 months and finds its maximum extent around the age of 5 months.

In infancy, different manifestations of communicative musicality meet the criteria to be called musical, including lullabies (Savage et al., 2021; Mehr et al., 2019; Bainbridge et al., 2021) and infant-directed singing (Trehub et al., 1997, Trehub et al., 2013, Trehub & Gudmundsdottir, 2015, Shifres, 2011, Corbell et al., 2013, Corbeil et al., 2016, Weiss et al., 2012). On the other hand, other manifestations of communicative musicality are not musical but do show antecedents of the temporal arts, granting them a musicality trait (Dissanayake, 2000a, b; Trehub & Cirelli, 2018). Three of the latter manifestations have been pointed out as particularly relevant in socio-cognitive development: *protoconversations*, *infant-directed improvised performances* (hereafter IDI-performances), and *action song*s. The three of them constitute forms that emerge in dyadic interactions, where adult actions are organized to accommodate infant behavior, in order to foster inter-subjective bonding. For this reason, we call them *emergent organizations of early communicative musicality.* Protoconversations organize in a behavioral alternation where adult and baby take turns to participate using facial expressions, vocalizations, and glances, while, on the other hand, infant-directed performances are linked to phrase sequentiality emerging through the repetition-variation form of short and simple sound-kinetic motifs; finally, action songs involve a culturally regulated combination of verbal phrases, sound, movement, facial expressions, glances, and touches. All three share the common feature of being playful activities that summon the baby to social interaction. The three emerging organizations exhibit another distinctive feature: they are often multimodal. As we shall see, even protoconversations involve, in addition to vocalizations, further exchange modalities. Certainly, other events equally committed to musicality are included in daily routines and serve other functions, such as lullabies or infant-directed nonverbal singing – humming – aimed to induce infant sleep, but we will not deal with them here.

These three emergent organizations of early communicative musicality have been explored independently. Among others, Bateson (1975), Bullowa (1979), Trevarthen (1977, 1998), Jaffe et al. (2001), and Devouche and Gratier (2019) studied protoconversations. Stern (1974), Beebe and Gerstman (1980), and Español and

Shifres (2015) conducted research on the notion of sound-kinetic performances (see also Chap. 2). On their part, Eckerdal and Merker (2009) approached the study of action songs in a systematic way. The three emerging organizations, with their specific features, shape the primordial adult–infant interactions and promote the infant incorporation of behavioral patterns essential for early subjective and intersubjective development and later human activity. Turn alternation in protoconversational exchanges provides a dialogue format between competent speakers and constitutes the basis of a collaboration that underlies the cooperation principle of Grice (1989). The repetition-variation form characteristic of performances supports the composition of phrases in the temporal arts. Normativity in the sound-kinetic coupling of action songs is inherent to ritual-like behaviors that tend to be socially shared and passed from one generation to the next. The three emergent organizations show up as variations of the same and unique intersubjective behavioral matrix linked to multimodal ludic activities. To the best of our knowledge, nonetheless, they have not been studied jointly. We lack information about the ontogenetic development of the three organization types integrating this matrix and the link between them. As in Chap. 2, our main interest lies in IDI-performances. In this chapter, we intend to provide new evidence for this matrix of multimodal interactive behaviors, focusing on IDI-performances and comparing their trajectory with that of other two emerging organizations.

In the 1970s, developmental research established that certain adult–infant behaviors could be understood as formally conversational even in the absence of verbalization (Bateson, 1975). The parties engage in turn-taking exchanges. The mother utters short phrases, words, interjections, and the infant replies with gurgles, smiles, murmurs or sounds, facial expressions, and/or movements. The similarity between the format of these primary dialogues and that of adult conversation led almost naturally to the acceptance of the term protoconversations to refer to these exchanges where mother and infant collaborate in a somewhat successful pattern of alternation and jointly produce a brief conversation-like performance (Bateson, 1979). Sometimes, when the infant does not respond vocally, the mother acts as if he or she has done so: after vocalizing, she will wait for a prolonged time (seemingly the sum of the average adult pause time in a dialogue [0.6 sec] plus the duration of an infant imagined vocal response [0.43 sec], plus again the average adult pause time in a dialogue) before speaking again (Beebe, 1982). While its salient component is vocalization, protoconversations also include other modalities. Sometimes the immaturity of motility only allows the infant to perform, in an interactive situation, an unsuccessful orofacial movement corresponding to a given vocalization, like attempting an unachieved utterance (proto-speech – Trevarthen, 1977). They have been extensively studied from their first formulation to the present day, and a certain general agreement has been established regarding their fundamental role in the development of communicative skills in general and language acquisition in particular (Bullowa, 1979, Fogel et al., 1988, Ginsburg & Kilbourne, 1988, Bråten, 2009, Demuth, 2013, Gratier et al., 2015, Hilbrink et al., 2015, Kokkinaki, 2010, Malloch & Trevarthen, 2009 – for an integrative report, see Devouche & Gratier, 2019).

Despite its acceptance, the term protoconversation has been used in different ways. Respecting its original description, some have delimited the phenomenon as an event composed of alternating vocalic turn cycles (Bateson, 1971, 1975, 1979, Beebe et al., 1985, Bloom, 1977, Jaffe et al., 2001; Lavelli & Fogel, 2013, Stern et al., 1975, Trevarthen, 1979; Stern & Gibbon, 1979). Other authors have used the category more broadly, encompassing face-to-face interactions in general (Delack & Fowlow, 1978; Demuth, 2013; Goldberg, 1977; LeVine et al., 1996; Schieffelin & Ochs, 1995; Rochat et al., 1999). Therefore, there is no single, precise and consensual definition. On the other hand, there is a tendency to focus on oral output. From our particular point of view, already defined as multimodal (see Chap. 2), and given the existence of data supporting a protoconversational structure unaided by vocalizations, we consider the adoption of a broad-spectrum definition to be appropriate. We believe that, to the extent that the intervention of the infant takes place within the interaction during his turn-taking, his behavior is driven by a willingness to engage in protoconversation, even if he does not vocalize. We also consider relevant reserving the term protoconversation for turn-taking oriented exchanges (successful or unsuccessful) rather than extending it to encompass other types of face-to-face interactions.

While it may seem obvious that being exchanges, protoconversations should be considered bidirectional, some authors (Kaye, 1982; Snow & Ferguson, 1977) claim they are one-way phenomena, since it is exclusively adults who establish time-space frames that encourage infants to join the interaction less randomly. Intervention-regulated alternation would then be a gradual achievement stemming from adult scaffolding (Hsu & Fogel, 2003). Accordingly, the idea of a true protoconversation format in early infancy would be questionable and infants would not be co-protagonists of the phenomenon at such an early stage as previously assumed. However, the bidirectionality option is supported by research on the early uptake of interactive contingency. Infant sensitivity to contingency clearly emerges by the age of 2 months (Striano et al., 2005). From that point on, infants react when social interaction is interrupted (Tronick et al., 1978) or if its temporality is artificially modified (Murray & Trevarthen, 1985, 1986), supporting the idea of a foundational bidirectionality and early adjustment rhythms between adult and infant behavior (Beebe et al., 2012). Both positions could reconcile by means of a protoconversational study that goes beyond primary intersubjectivity (Hilbrink et al., 2015) in which two relevant changes were identified in this interaction type. The first, at around 5 months, indicates an overlapping decline in turn-taking and therefore an increasing alternation achievement. The second, at around 9 months, shows some deceleration of infant interventions, taking longer pauses to respond (maternal pauses retain their length). The enhanced synchronization at 5 months may stem from a better understanding of protoconversational play that began when the adult was still unilaterally providing the infant with the relevant scaffolding to engage. The change around 9 months could be explained by the progressive inclusion of worldly objects in the exchanges, as a result of joint attention. Therefore, protoconversations would be subject to alterations derived from the interweaving of different forms of cognitive support that initiate psychological life.

As Carretero and Español (2016) pointed out, in the 1970s, adult–infant interaction research – for example, the studies of Bateson (1971, 1979) on protoconversations and the work of Condon and Sander (1974) on interactional synchrony – were sparse compared to the total body of research on infancy and early childhood. Such studies that focused on the adult–baby dyad were influenced by three academic fields: nonverbal communication, ethology, and adult verbal communication. Her training with anthropologist Ray Birdwhistell, the originator of kinesics, allowed Bateson to appreciate that in protoconversational turn-taking, vocalizations were accompanied by mutual eye contact and facial response. Along the same lines, Trevarthen (1977, 1998) described protoconversations as exchanges in which infants focus intensely on the face of the mother and then actively smile, followed by a set of expressive behaviors to which the mother responds with vocal textures, facial expressions, and specific postural actions, occasionally touching the infant. Protoconversations are an exchange mode where protagonists orient themselves to each other in an effort to share emotional and behavioral communion. Although vocalizations predominate, other behaviors such as caresses, facial expressions, and postural changes are also part of the exchange. Finally, it should be noted that although pioneering studies refer to protoconversations in the first months of life, this phenomenon is being reported in neonates (Boiteau et al., 2021; Dominguez et al., 2016) and preterm infants (Kiepura et al., 2022; Oller et al., 2019).

Whereas studies of protoconversation were driven by the linguistic turn in the twentieth century, the description of adult–infant dyadic performances is a product of the corporeal turn, characteristic of the twenty-first century (Sheets-Johnstone, 2015). However, it is important to note that the discovery of these interactive performances was only possible thanks to the pioneering studies of the last century that focused on interactive movement between adults and infants. Stern (1974; Stern et al., 1975) and Beebe and Gerstman (1980) highlighted that maternal performances toward infants consist of multimodal behavioral reiterations that compose sound-kinetic phrases. Beebe and Gerstman (1980) referred to these as multimodal package behaviors. However, this line of research was not extended, and it left movement studies behind to refocus on vocalizations (Beebe, 2014). On the other hand, research on infant self-development within dyadic emotional interactions with adults described early exchanges as mutually co-regulated movement interactions and as multifaceted and complex communicative processes exceeding the notions of conversation and language (Fogel & Garvey, 2007; Garvey & Fogel, 2007). The reciprocity of a joint know-how builds base structures on which the associated variations are subsequently developed (Lyra, 2007). But IDI-performances, as we understand them, were very gradually defined following studies that highlighted the antecedents of temporal arts (music and dance) in early interactions (Dissanayake, 2000a, b) as well as movement as the axis of psychological and social life (Stern, 2010). Following this last formulation by Stern, Español (2004) acknowledged IDI-performances as temporal arts-related specific interactive phenomena, different from protoconversations and other interactive formats in two central aspects: their particular organization – the repetition-variation of simple motifs – and its particular variations – made with resources specific to the temporal

arts. Further microanalytic studies were conducted from then on to narrow its definition (Español & Shifres, 2015; Ospina Tascón & Español, 2014; Español et al., 2015). Its last formulation can be found in Chap. 2 of this volume. Unlike what happens in protoconversations, the adult organizes his performance for and on behalf of the infant using the repetition-variation form. This particular interaction does not focus on the alternation of turns, but rather on phrase sequence. The performances reflect an adult impulse to improvise short phrases consisting of a beginning, a development, and an end. As a starting point for their improvisation, adults take a movement made by the infant, or a sound he/she can vocalize, or both, or an environmental incident (noise, lighting change), or their own bodily sensations, or past memories brought into the present interaction. They thus form motifs, sound-kinetic units a few seconds long that are repeated, resulting in short phrases. Motifs and their repetitions lead to phrases constituting IDI-performances. The repetition-variation form is characteristic of the temporal arts (music and dance). So are the resources used by adults to vary the motif in their repetitions – such as the variation of melodic contours, timbre, and intensity for sounds; shape and effort for movements; or speed and rhythm for sounds and movements. Therefore, IDI-performances are structurally and functionally connected to the temporal arts. See Chap. 2 for a detailed presentation of IDI-performances.

The last type of these emergent organizations is called action songs, defined by Eckerdal and Merker (2009) as a composition of music, lyrics, and mimicry, codified by each culture and transmitted from parents to children. These songs are associated with a prefixed and relatively unchanging scheme of actions and movement, enduring across generations. This emerging organization involves word, movement, and melody units consubstantiated and combined in a narrative sequence. The action song is performed in nurturing and/or educational contexts. It is part of our original cultural reservoir and, albeit tolerant of performance variations, preserves its fundamental nucleus largely unaltered. A well-known example in Spanish-speaking countries is an action song called *Linda manito (Cute little hand)*; a close English-speaking example is an action song called *Itsy Bitsy Spider*. Unlike *Itsy Bitsy Spider*, which plays with finger movements, *Linda manito* involves opening, closing, turning, hiding, and showing hands. Adults perform the song in front of the infant, setting up a multimodal scaffold for him/her to reproduce sounds and movements following the melody and rhythm of the song. This is a conventional and clearly regulated way to link sounds and movements, summoning observation, repetition, and transmission. Sometimes actions represent what the lyrics say, sometimes they do not, and the link between the movements and the words or sounds is arbitrary, but in all cases the pattern of coreations is prescribed by culturally established implicit or explicit norms . Initially, infants participate by drawing from their preverbal communicative background, but action songs are oriented to progressively incorporate both action and vocalization patterns that comprise them. Eckerdal and Merker (2009) propose action songs as a means of introducing infants to ritual co-participation routines, a gateway to communal practices where, as in rituals, structural rules lacking efficacy in practical life matters must be respected and are only valid within the limits of the action song. Action songs embody the

broad scope of mimesis unfolding in any ritual activity. It invites infants to try to imitate what they see and hear, facilitated by the connections between songs and associated actions. Unlike protoconversations and IDI-performances, action songs (not by this name) have been collected, interpreted, and studied in different areas, as they are part of widely recognized parenting play (Calmels, 2014).

While IDI-performances share musicality and movement display with action songs, they lack the iconographic-kinetic illustration of texts (lyrics) frequently found in them, as well as the prefixed sonorous-kinetic patterns that characterize them when iconographic illustrations are not evident. Both IDI performances and action songs move away from mere singing or sound stimuli, but in IDI-performances there is an absence of the semantic mimicry prescribed by the lyrics that make up the song, and an adherence to normative ways of coupling sound and movement. Improvisation defines IDI-performances just like normativity defines action songs. In Chap. 2, we made a distinction between ritualized IDI-performances and totally improvised IDI-performances. Ritualized IDI-performances are based on the cultural repertoire of play; however – as shown in the microanalysis of a peek-a-boo play event in Chap. 2, Sect. 3.1, Moment 2 – adult improvisation modifies, transforms, and aesthetically elaborates these culturally dependent structures. Ritualized IDI-performances are somewhere in between IDI-performances and action songs: improvisation brings them closer to the former, and cultural repertoire belongs to the latter. As we will see in the observation code, in this chapter we incorporate ritualized IDI-performances into the action song category. This decision was made because determining their level of improvisation would require a detailed analysis beyond the scope of this chapter, while their regulatory nature, which brings them closer to action songs, is evident.

The three emergent organizations described have been explicitly linked to the concept of communicative musicality. Malloch and Trevarthen (2009) present a protoconversation between a 6-week-old baby and her mother as a prototypical example of communicative musicality. A spectrogram of the vocalic exchange between them reveals an asymmetric interaction: we observe the participation of the baby but also how it differs from that of the mother, who responds with several complex utterances to the short and scarce vocalizations of the baby. An analysis of the timbre component (an aspect of quality) shows a repeated and particular maternal vocalic imitation: after each baby utterance, the three timbre values corresponding to the mother decrease. The mother renders a voice more similar to that of the baby. They suggest that by using this single resource, the mother is signaling her baby that she is there, attentive, and ready for an exchange. To describe infant-directed performances, I. Martínez (2014) explicitly incorporated the concept of compositionality, typical of music composition studies. She suggested adults possess an intuitive compositional ability that allows them to arrange musical expressive materials by manipulating temporality.

In this arrangement, redundancy across the various expressive modalities – sounds and movements – help the repeated motif to gain structural coherence. Español and Shifres (2015) analyzed infant-directed performances using methods from the temporal arts (the Laban system of movement analysis and the expressive

analysis of music performance); they explicitly treated them as communicative musicality phenomena finding structural and expressive aspects compatible with the formal structures and expressive resources of Western music and dance. I. Martínez et al. (2018) showed the three dimensions of communicative musicality in performances, making particular reference to the image-schemas and primary metaphors they contain. These topics are fully addressed in Chap. 2 in this volume. Finally, Eckerdal and Merker (2009) presented action songs as another prototypical case of infant communicative musicality and analyzed their three constitutive dimensions: pulse, quality, and narrative.

3.2 An Empirical Study on Infant-Directed Improvised Performances, Protoconversations, and Action Songs

In this chapter, we aim for an integrated study explaining how these three emergent organizations of communicative musicality perform between the ages of 2 and 9 months, a time span that includes primary intersubjectivity and the transition to secondary intersubjectivity. We conducted a longitudinal case study of a mother–infant interaction with the overall purpose of analyzing them in their ecological context.

3.2.1 Method

Participants and Materials

We observed a mother–baby dyad from a middle-class Argentinean family residing in the city of Buenos Aires. They were selected among people close to the first author of this study. The baby was the second child and showed no alterations or illnesses during development. He was aged 00; 2 (20) in the first session, and 00; 9 (26) in the last session, according to the Piagetian convention [year; month (day)]. During the observations, the dyad had the usual objects of their daily life at hand (toys, household objects offered by the mother). A SONY DCR-SR82 camera was used to film the sessions.

Procedure

The dyad was filmed in their home every 15 days performing 30-minute sessions. The mother was instructed to behave and play with her baby as she normally did. The sessions included all the activities taking place in the parenting routine – feeding, diaper changing, and social play – and were held in different locations throughout the home. The sessions, which totaled 15, were filmed by the first author of this

paper. At the end of each session, the first author asked the mother certain questions that helped understanding some aspects of the interaction, especially those that were part of the shared history of the dyad, allowing to complement the registered material with daily out-of-session occurrences. Through these questions, we were able to determine which performances had already been carried out in other moments, before or after they first appeared in the sessions.

We analyzed the filmed material using the Constant Comparative Method (Glaser & Strauss, 1967; Strauss & Corbin, 1994), through which we adjusted the observation categories in several successive phases in line with the different aspects of the subject phenomenon. We used the Anvil 5.1.9 video annotation software to analyze the videos.

During the first phase, seeking to establish clear nominal and operational definitions, we watched the entire footage and identified those events where any of the above-mentioned emergent organizations could be recognized. We then produced observational descriptions of all the events from five sessions of different developmental moments (months 3, 5, 7, and 9). Concerning protoconversations, we found that the elements normally accounted for (sounds or mouth movements, glances, and facial expressions between adult and infant) were generally complemented by body contact or touches also part of the mutual involvement. Regarding performances, we found that sound, movement, facial expression, gaze, and touch phrases were organized as repetition-variations of simple motifs. As discussed in Chap. 2, the repetition-variation of a sound and/or movement motif configures a structure of meaning, a unit with a beginning, development, and end. Through repetitions, motifs spontaneously organize themselves into holistic phrases; that is, they form discrete units within the uninterrupted interaction flow. We observed motif variations in the specific dimensions outlined in Chap. 2: spatial, temporal, and/or energetic dimensions.

Contrary to expectations, given the fleeting nature of the performances, a second phase showed that some were repeated. We distinguished between fleeting performances, which only appear once, take on a clear form, and are never repeated; and others that, on the contrary, consolidate within the exchanges and replicate more than once, which we called reiterated performances. After isolating the recurrent IDI-performances, we asked the mother if they were spontaneously arising inventions generated during the interaction, or if they belonged to the social play repertoire of the family, transmitted by parents and grandparents. We subdivided the reiterated performances according to the responses given by the mother into those spontaneously generated in dyadic interactions (therefore called dyadic) and those that were part of the adult–infant play traditions in the family (therefore called familiar). The category of reiterated performances was inspired by the work of Lyra and Valsiner (2011) on historical dyadic interactions, intersubjective engagement styles generated in the bonding-interactive history of each particular dyad and repeated on a simple script.

Observation Code

The final code was configured into the following observation categories:

IDI-performance: It is a short multimodal spontaneously constructed phrase by adults for their infants, based on the repetition-variation form of minor units or expressively modeled sound-kinetic motifs. Simple motifs must be reiterated at least twice. Sound and movement modeling is spatial, energetic, and temporal. Spatial modeling includes melodic contours, timbre variations, and variations in movement shape: rising-descending, spreading-enclosing, and advancing-retiring. Energy modeling includes relative sound intensity degrees and movement effort variation. Temporal modeling refers to speed variations, tempo variation and duration, and time and rhythm for sound and movement (for a detailed description of spatial, energetic, and temporal modeling, see Chap. 2).

Fleeting IDI-performance: An IDI-performance that only occurs once during the developmental period studied.

Reiterated IDI-performance: IDI-performance that appears more than once during the developmental period studied. It is divided into two types: dyadic reiterated performance, repeating a motif created during dyadic interactions, and familiar reiterated performance, including a motif coming from the repertoire of the family (although the adult usually modifies the structure of multimodality and prefixed musicality during the repetitions).

Protoconversation: Playful social interaction where adult and infant take turns to exchange sound utterances, gaze patterns, and facial gestures as in a real dialogue situation (except that in this situation the elements of the interaction are not necessarily semantic). The vocalic intervention of the infant may be an actually uttered vocalization or soundless lip movements (proto-speech). There should be at least two turn-taking cycles.

Action song: It is a song accompanied by ritualized mimicry that iconically illustrates the content of the lyrics or where pre-established movements coupled to the song are performed in a normative way. The ritualized performances described in Chap. 2 are included in this category. These – we recall – are regulated ways of combining sounds and movements coming from cultural play repertoire, but which also involve high levels of adult improvisation. Since the aim of this study was not to discuss the improvisation extent of every single play stemming from the cultural repertoire, and at the same time considering that ritualized performances carry some degree of pre-established normativity, we included them in this category. For this reason, and in the context of this study, IDI-performances are understood as the original IDI-performances in the previous chapter.

Two coders (the first two authors of this chapter) participated in the coding process, followed by a discussion where there was disagreement. In the few instances where no consensus was reached, the event was dismissed. This procedure is consistent with that used in social interaction studies (Forster & Iacono, 2013, Jonsson et al., 2001, Español et al., 2015, and Chap. 4 in this volume).

3.2.2 Results

We first offer a quantitative analysis, starting with the general data collected from the three emerging organizations and then focusing on IDI-performances. Second, we provide a qualitative analysis of the elements of musicality (sound and movement) in each type of IDI-performances.

Quantitative Analysis

The data obtained from what is, to our knowledge, the first joint empirical study on the three emerging organizations of communicative musicality are presented below. Table 3.1 shows the frequencies of the three emerging organizations in each recorded session.

The highest frequency corresponded to IDI-performances (54.7%), followed by action songs (35.1%) and protoconversations (10.1%). IDI-performances and protoconversations were observed from the beginning of our research (second month of life), while action songs emerged by the end of the fourth month. In the middle of the sixth month, precisely when IDI performances reached their maximum frequency, protoconversations abruptly and definitively dropped to zero, after having reached their maximum frequency in the previous session (session 8). Action songs reached their maximum frequency in the eighth month. IDI performances were the only ones present during the entire period studied.

Table 3.2 shows the frequencies of the different types of IDI-performances throughout the sessions, according to the fleeting/reiterated criterion and differentiating, within this second class, between dyadic IDI-performances, idiosyncratic products arising from the mother–baby interaction, and family-related IDI-performances, pertaining to the family games of the mother, who transferred them to the exchanges with her son.

Out of the total number of IDI-performances (81 occasions), 41.9% were fleeting (34 occasions), and 58% were repeated (47 occasions). The frequency of fleeting IDI-performances declined and almost disappeared after session 11 (7 months). The frequency of reiterated IDI-performances remained erratic throughout the period. In the first 3 sessions, the frequency of fleeting IDI-performances was higher than that of reiterated IDI-performances but from the fourth session on (4 months) the frequency of reiterated IDI-performances was equal to or higher than that of fleeting IDI-performances, with one exception in session 10 (7 months). The first dyadic reiterated performances occurred at 4 months, being the latest to appear (the familiar reiterated performances emerged in the second session). Interestingly, fleeting and dyadic reiterated performances, both spontaneously organized by the adult, were, respectively, the first and the last variant in this type of emerging organization.

Table 3.3 shows the breakdown of reiterated, dyadic, and familiar IDIs, identifying the different varieties registered within each of these two groups. These were

Table 3.1 Frequency of emerging organizations in the period under study

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Age (in months)	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	
IDI-performances	4	9	6	5	8	9	7	4	10	4	6	1	2	2	4	81
Proto-conversations	2	0	2	0	4	1	1	5	0	0	0	0	0	0	0	15
Action songs	0	0	0	0	2	2	8	3	0	5	5	8	12	5	2	52
Total	6	9	8	5	14	12	16	12	10	9	11	9	14	7	6	148

Table 3.2 IDI-performance types according to sessions

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Age (in months)	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	
Fleeting	4	8	4	0	4	4	2	2	2	3	0	0	1	0	0	34
Reiterated	0	1	2	5	4	5	5	2	8	1	6	1	1	2	4	47
— Dyadic	0	0	0	5	0	0	3	1	5	0	4	0	0	0	1	19
— Familiar	0	1	2	0	4	5	2	1	3	1	2	1	1	2	3	28
Total	4	9	6	5	8	9	7	4	10	4	6	1	2	2	4	81

Table 3.3 Frequency evolution of reiterated performances

Session	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Age (in months)	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	
Dyadic	0	0	0	5	0	0	3	1	5	0	4	0	0	0	1	19
Little ants	0	0	0	1	0	0	1	0	2	0	0	0	0	0	0	4
Gaze interplay	0	0	0	2	0	0	1	1	0	0	3	0	0	0	0	7
Kisses and percussive lip contact	0	0	0	2	0	0	1	0	3	0	1	0	0	0	1	8
Familiar	0	1	2	0	4	5	2	1	3	1	2	1	1	2	3	28
Barambuch	0	0	1	0	1	1	0	1	0	0	0	1	1	1	0	7
Stinguishen	0	1	0	0	2	2	1	0	0	0	0	0	0	0	0	6
Smelling paw	0	0	1	0	1	2	0	0	0	0	1	0	0	0	0	5
Seesaw	0	0	0	0	0	0	0	0	2	1	1	0	0	1	3	8
Approaching an object with sound	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	2
Total	0	1	2	5	4	5	5	2	8	1	6	1	1	2	4	47

named (i) based on their outstanding phonetic motif ("Barambuch," "Stinguishen" – expressions without meaning, approximate phonetizations), (ii) based on a brief description of the actions/movements involved in the exchange (seesaw, kisses and percussive lip contact, gaze interplay, approaching an object with sound), or (iii) by a verbalization that centralized the mother–baby back and forth ("*Hormiguitas*"/ little ants, "*Olor a pata*"/smelling paw).[1] The qualitative analysis (Sect. 3.2.1.2) provides a prototypical observational narrative for each of them, illustrating the type of play in detail.

The most frequent reiterated IDI-performances were *Kisses and percussive lip contact* and *Seesaw* (8 occasions each), followed by *Gaze interplay* and *Barambuch* (7 occasions each). Thus, the two highest frequencies are represented in both groups of reiterated performances. This indicates the absence of a hegemonic type of performance between both groups, dyadic and familiar: the adult chose both her own sound-kinetic structures and those of her family repertoire.

[1] Since the names of these types of reiterated IDI-performances correspond, in some cases, to actual vocal emissions, we have used the first time the relevant double quotetion marks (e.g. "Barambuch"), but from now on they will be systematically italicized (e.g.Barambuch) to match them with other types of emergent organizzations.

We registered eight different action songs (each one will be briefly described in Sect. 3.2.3). Table 3.4 details the frequency of action songs by age and session. The most frequent action songs were *Linda manito/Cute little hand* (present in 13 occasions), *Tortita de Manteca/Butter pancakes* (11 occasions), *Caballito gris/Little gray horse* (8 occasions), and *Conga* (the name of a dance song) (7 occasions). *Pimpón el muñeco/Pimpón the doll, ¿Dónde está?/Where is it?* and *Beso esquimal/Eskimo kiss* were the least frequent.[2] The first action song to appear in the studied period (around the fourth month) was *Cute little hand*. *Tachin-tachin* (invented expression) was the last to emerge around the age of 9 months, during the penultimate session. In sessions 12 and 13 (at 8 months), we found the highest frequency and maximum diversification of action songs, with four varieties in each case.

We measured the duration of the three types of emergent organizations in seconds and compared the mean durations. While the differences were not statistically significant ($F[194–2] = 2.367$; $p = 0.097$), protoconversations were the longest lasting emergent organizations on average (20.8 seconds), followed by IDI-performances (17.2 seconds) and action songs (14.6 seconds).

Qualitative Analysis

Next, we provide observational accounts for each type and subtype of early interactive musicality emergent organizations found in the study, beginning with IDI-performances, fleeting and reiterated, followed by protoconversations, and ending

Table 3.4 Action song frequency by age and session

Sessions	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	Total
Ages (in months)	2	3	3	4	4	5	5	6	6	7	7	8	8	9	9	
Little gray horse	0	0	0	0	0	2	5	0	0	0	0	0	0	0	1	8
Cute little hand	0	0	0	0	2	0	2	1	0	2	0	0	6	0	0	13
Pimpón the doll	0	0	0	0	0	0	1	2	0	2	0	1	0	0	0	6
Butter pancakes	0	0	0	0	0	0	0	0	0	0	5	0	4	2	0	11
Conga	0	0	0	0	0	0	0	0	0	1	0	5	1	0	0	7
Where is it?	0	0	0	0	0	0	0	0	0	0	0	1	1	0	0	2
Tachín-tachín	0	0	0	0	0	0	0	0	0	0	0	0	0	3	0	3
Eskimo kiss	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	2
Total	0	0	0	0	2	2	8	3	0	5	5	8	12	5	2	52

[2] As with reiterated performances, hereafter we will use italics and the English version, thus indicating that they are types of action songs.

with action songs. Subsequently, we analyze the most salient features of early musicality within them.

IDI-Performances

Fleeting IDI-Performances

Example A – Age: 0; 4 (9). Duration: 26 seconds.

The mother is sitting in an armchair. The baby is sitting on her lap, held by the armpits. She lifts him up and very slowly pulls him toward her, smiling and saying, "Co...go...te" (Spanish colloquial for neck). The movement of pulling the baby closer to her coincides lengthwise with the vocalization and culminates when the mother uses her lips to make contact with the neck and cheek of the baby. The mother then gives the baby a series of rhythmically spaced kisses in that area. She then returns the baby to his original place, on her lap, and repeats the whole action: she lifts the baby and pulls him toward her, vocalizing at the same time: "Co...go...te...te...." About halfway through, anticipating the outcome, the baby laughs and she hastens the kisses. This motif, based on a scheme of bringing their bodies together while vocalizing (segmenting the movement with a double intersyllabic pause), is repeated about three more times (four total occasions, with changes in speed, i.e., in the behavioral temporality).

Example B – Age: 0; 4 (28). Duration: 29 seconds.

The mother is seated holding the baby in her arms, lying down, facing her. As she takes one hand of her baby, she brings it to her mouth and begins to move it between her lips while vocalizing one sound: "Baaaaa..." Her vocalization is as long as the movements she inflicts on his hand. When she is finished, she pauses and smilingly observes her baby for a few seconds. In turn, he looks at her also laughing. The mother repeats the motif (moving the small hand of the baby back and forth over her own lips adding the utterance "Baaaaa...") five more times, varying the tone (spatial dimension, which includes the tone spectrum), duration, and rhythm (temporal dimension), for sound and movement.

Example C – Age: 0; 9 (11). Duration: 17 seconds.

The baby is sitting at a table facing his mother. She picks up a spherical toy and taps it against her free palm. She then extends that palm out in front of the baby, revealing there is nothing in it and hides the toy between both hands. He opens and closes them twice in a pulsating manner. Each opening and closing movement matches both beats in a vocalization: "Tu...tu...." She stands still, waiting for his reaction. Effectively, he begins to bring his hand closer to hers, reaching for the toy. She then performs the same opening and closing movement again, slowly, with the same vocalization: "Tu...tu...." The baby reaches out to grasp the hand of his mother and tries to open it. She gives in, opens her hands, and loudly exclaims: "Yes, he's got it!" The performance ends with the baby grasping the object and the mother celebrating, varying the celebrating tone (spectral dimension) with each repetition. This motif is repeated three times altering the speed (temporal dimension) and intensity (energetic dimension).

The three previous examples reveal a variety of movement and sound coupling intuitively produced by the adult from implicit patterns of early musicality. In A, the approaching movement is supported by a series of sounds coinciding with a verbalization between the beginning and the end, naming the part of the body where the whole performance ends, the neck. In this case, there is a peculiar synchronicity between sound and movement: the word that co-extends with the movement designates the zone where the motif ends. The movement of bringing the baby closer to

make the mouth and neck-cheek contact is performed in three syllabic beats, resulting in a touch that leads to a cluster of exploding sensations: melodic contour, the experience of being displaced (movement of the self -given by someone else), various kinds of touches, and finally kisses that are at once friction and a rubbing in the folds of the neck. There is also a special tone of intimacy. An *accelerando* brings the performance to an end, marked by a *finale* typical of musical compositions, with a full orchestral outburst that maximizes the intensity. This motif, consisting of small vocalizations that support a reasonably extended movement, contrasts with example B, where the mother conversely interweaves repeated movements with a long vocalization. While in B the rubbing between lips and hand coincides with the beginning and end of "Baaaaa..." demarcating the motif between both modalities, in A there are three syllables and only one movement. In any case, if IDI-performance A contains an action divided by the syllables of a word simultaneously uttered along that action, and if, on the contrary, this pattern is inverted in B because the vocalization harmonizes with brief oscillation movements, both differ from the multimodal organization in example C. Here the temporal coincidence between the opening–closing hand movements and the short vocalizations is perfect. The sound scansions correspond with the opening–closing on a one-to-one basis. Each motif movement comprises a set of sound and kinesis where both parts (and modalities) match perfectly. To put it another way, this performance consists of small unequivocal bimodal blocks, sharply separated by an exact temporal convergence. The "Tu...tu..." syllabication pattern and the playful movement of showing and hiding the object between the hands of the mother offer a third cooperation alternative between modalities. On the other hand, while the motifs in A and B represented a continuum with different points of departure and arrival, this motif is made up of simple or discrete components reproduced successively, with a pendular itinerary, since each action is followed by its opposite (reveal/unveil, open/close). This reversibility does not require phonemic differentiation but uses the same syllable to mark the pulses. In the three cases, A, B, and C, coordination marks a unity: sound and movement mutually reinforce each other to stand out from the background of the other previous and subsequent actions and to show themselves possessing a specific and independent behavioral value (sound-kinetic phrase).

The above examples show three different collaboration varieties between sound and movement according to the principle of communicative musicality. These varieties show how different possibilities of temporal scansion are elaborated in early playful interaction. In fact, the fundamentally dyadic format of musicality facilitates the management of temporal flows in infancy beyond those established by everyday life activities. Under this playful form, temporality can become a plastic, flexible matter, unfolding before infant sensitivity and cognition. In this way, a first taste of freedom of action and thought insinuates into human life. The oscillating movement that takes place when we meet others reveals that temporality is a new dimension fluctuating at every step. In their spontaneous performances, through repetition and variation, adults show babies that experiences can be contingently altered, return as repetition, and derive in ever new forms. In this case, novelty originates in the interplay between the body and the temporality of the other, in a paired motricity, in

patterns that settle down as the foundational psychological experiences of any form of intersubjectivity.

By definition performances are always simple strategies, therefore accessible to the baby. Perhaps because of such simplicity, this phenomenon has been overlooked until recently. This is especially understandable in the case of fleeting performances, since they never recur, which makes their detection even more difficult.

Reiterated IDI-Performances

Sometimes IDI-performances are replicated and remain in the repertoire of the dyad for a period of time. To the extent that they are repeated on different occasions (see Table 3.3), any indication of age, duration, or session is omitted below.

Reiterated Dyadic IDI-Performances

Little ants. The adult uses her fingers (finger walk) or an object to walk over the body of the baby, following a tactile stimulation trajectory matching the syllable segmentation of the words *li-ttle-ants* in every step. Variations fall on the finger movement speed and syllable pronunciation (temporal dimension), and on the finger movement amplitude and the generally ascending tonal contour (spatial dimension). This performance was observed on 4 occasions and 3 sessions out of 15 total sessions.

Gaze interplay. The mother looks away from the baby and while holding an averted gaze vocalizes "Ahhh." She then proceeds to look for the gaze of the baby. The interaction plays with opposites "I don't look at you pretending to be distracted, I look at you and smile." While the mother keeps her eyes off the baby and vocalizes "Ahhh," she remains impassive, still, generating expectation, until she suddenly turns her head toward the baby looking for his gaze. The mother plays on the expectation of anticipation inspired in the baby by knowing that her gaze will return at any moment. The variations occur primarily in the duration of visual contact moments and separation: a dramatic tension surrounds the exact moment she will look at the baby again. This performance was observed on 7 occasions and 4 sessions out of 15 total sessions.

Kisses and percussive lip contact. The mother makes a series of puffs on different parts of the body of her baby. Bringing her mouth close to make skin contact, she makes sounds by blowing and puffing. The puffing involves bimodal sensations of sound and tactile vibration in conjunction with adult lip contact. The mother consistently puffs three times over the body of the baby in such a way that each series traces a different path. She varies the rhythms in each puffing series and alters the intensity, duration, and speed of the contacts with each repetition. During some of the series, the puffing action became so soft that it was better described as kissing. This performance was observed on 8 occasions and 5 sessions out of 15 total sessions.

Little ants can be placed in line with example C of fleeting performances given that each syllable emphasizes each finger step on the body of the baby. *Gaze interplay* is different because even if movements are orally reinforced by the mother (informative redundancy), she refrained from repeating the same vocalization or offering a contrast of sound elements, and instead clearly separated movements from each other using silence and vocalizations. Vocalizations disappear to enhance the contrast between sensing and not sensing the gaze of the other, between being the focus of attention and being ignored (needless to mention, the impact of this on the psyche

of the infant). *Gaze interplay* could also be placed, like *Little ants*, in series with example C of fleeting performances. They both perfectly match movements and sounds (here, presence/absence of sound) and action reversibility, again movements following each other in the opposite direction (open/close, look/do not look).

Finally, the puffing and kissing performance over the skin of the baby – tactile-auditory stimuli – reflects once again the exact coincidence between both modalities. We should keep in mind in this case the relevant difference with the previous two types of dyadic reiterated IDI-performances, and with *Little ants* and *Gaze interplay* where the affected modalities, movement, and vocalization might not have been involved. The finger walk could have been performed unaided by syllabication, and the series of syllables could have stimulated the sensitivity of the baby unaccompanied by matching movements. Due to an implicit decision of the adult, IDI-performances usually include more than one modality. In this case, the kiss or snort generally involves simultaneous skin pressure and sound, a product of the natural relationship between lip morphology and their sound. But its performativity character emerges because the kiss or puff is repeated in a sequence, turning from mere kisses–puffs into a temporally ordered structure according to reiterated pulses, and also because these sounds were exaggerated and the contact with the body of the baby was made with unnecessary pressure and duration, as if these aspects were jointly elaborated for a ludic-musical stimulation. This transforms puffing and kissing into musicality made with puffing and kissing.

Reiterated Family IDI-Performances

Barambuch. The baby is sitting on the lap of the mother, facing her. She takes him by the armpits and pulls him closer to her, while emitting a playful vocalization that accelerates as their heads get closer together: “Baram, baram, baram...,” until they finally touch, and then the outcome is underlined by the exclamation "Buch!” Then the mother pushes him away, sitting him back on her lap, and the motif is repeated. Each “Baram” vocalization accompanies the approach and the final vocalization “Buch!” accompanies the culmination of the game or climax: both heads touching each other. This performance was observed on 7 occasions and 7 sessions out of the 15 total sessions.

Stinguishen (expression without translation). The adult slides her hands over the body of the baby, who is lying down. The mother places her hands on the shoulders of the baby while caressing him lengthwise in a wide trajectory down to the feet, exerting light pressure and vocalizing, as she goes over the little body, “Stinguishen, stinguishen, stinguishen....” This sequence is the motif, repeated several times thereafter. The baby is usually willing and stretches, gurgles, and smiles, as he is touched. This performance was observed on 6 occasions and 4 sessions out of 15 total sessions.

Smelling paw. The baby is facing the mother, lying on her legs. The mother takes the foot of the baby, brings it to her nose, smells it (or pretends to smell it), and, smiling, says to the baby, “What a smelly paw!” (with slight variations: “Oops, what a smell!” or “What a smelly dirty paw!” exaggerating the tone, higher pitched than normal) and then releases it. The baby usually looks and smiles at her, and on more than one occasion, he offers her his foot. The motif is constituted by the movement of grasping the foot, sniffing it, verbally expressing displeasure, adopting a corresponding facial expression, and finally releasing the foot, while looking at the baby. This performance was observed on 5 occasions and 5 sessions out of the 15 total sessions.

Approaching an object with sound. The mother invites the baby to interact by approaching an object and at the same time emitting a vocalization that temporally coincides with the movement to its full extent. Then she quickly withdraws the object. The mother emits a variety of sounds: prolonged sounds such as "Ahhh," a syllable repeated a certain number of times ("Pi, pi, pi, pi, pi...," "Chi, chi, chi, chi..."). The sound lasts the length of the movement of bringing the object closer. Variations are concentrated in the intensity, speed, and duration of movements and sounds. This performance was observed on 6 occasions and 3 sessions out of 15 total sessions.

Seesaw. Kneeling on the floor (but sometimes the same performance happens with the mother standing), the mother holds the baby in her arms and tilts him backward, bringing him upside down. After a pause, she brings him back to the normal position. Each kinesthetic moment of this motif (the head-down positioning, the pause, and the return to the normal head-up position) is accompanied by oral remarks. In the first and third moments, some vocalizations help emphasize the positional contrast. The second moment coincides with a moment of silence. From the third repetition onward, the baby actively participates, tilting back his head before the mother starts the movement (promoting the continuation of the motif repetitions). This scheme is mounted on the vertical axis of the body of the baby and plays with up-down opposites. It was observed on 7 occasions and 3 of the 15 sessions. This performance was observed on 6 occasions and 4 sessions out of 15 total sessions.

The first of these performances, *Barambuch*, integrates the approaching movement with the accelerated rhythm and a finale that involves, as in example A of the fleeting IDIs, a final vocalization twist. In that case, the change was less salient, in terms of the integral motive, because the series of sounds was composed of three distinct syllables, whereas in *Barambuch* there is an iteration of the first two until a substitution takes place at the end. Thus, the vocalization accentuates the climax of the entire motive, breaking the preceding gradual advance. The *Stinguishen* performance is also structured similarly to performance A in the fleeting class, albeit for a different motif. The same vocalization is repeated several times within the length of a single prolonged movement (respectively, the movement of the body of the baby with the hands of the adult and the distal-proximal movement). The sonority of *Smelling paw*, as opposed to the two previous cases, relies on the language (as in *Little ants*), suggesting that in the performative activity of early musicality, the oral content oscillates between words, syllables, phonemes, and noises regardless of their importance. In fact, words are words only for adults, while for infants they indistinctly accompany the set of movements, highlighting amodal properties through various combinations (see Chap. 8). *Smelling paw* is significant because it involves facial expressions. In any performance, particularly in those ending in a sound-kinetic outburst, the face of the adult is usually altered at the climax, and the different vocalization pulses or sequenced movements are usually reflected in the features, but in this case the adult uses a concrete and specific facial expression, linked to the manifestation of a basal emotion: disgust or aversion. While the baby may be able to identify, through a cross-modal correspondence, this expression and may express it involuntarily in his facial musculature, we must emphasize that here, the playful mood is reflected in the tone of the interactions, and that the disgust on the face of the mother intersects with smiles and playfulness. At the same time, this facial expression is voluntarily exaggerated, endorsing the construct of motionese, a bodily correlate of expressive affect that is orally manifested as motherese and

that, much like this one, is spontaneously used in interactions with infants (Brand et al., 2002; Brand & Shallcross, 2008). This attitude of adult expressivity, involving vocalizations, movement, and emotional expression, is called multimodal motherese (Gogate et al., 2000), foregrounding the functional structure of cross-modal solidarity in addressing infants and pregrammatical children. To summarize, *Smelling paw* is a unique convergence between adult movements (contacting the body of the baby and performing certain movement forms on it, like extending the foot to smell it and then releasing it), a facial expression linked to an innate emotion, a spontaneous enhancement in expressivity (motherese + motionese), and a certain thematically relevant verbalization, not mere vocalizations (indicating that insignificant sounds and language elements coexist in adult speech directed to infants and in the performances addressed to them). This conjugation of affectively convergent information is effective because it capitalizes on the possibilities for an early musicality in infancy, the first dialect for mutual understanding.

In the performance of *Approaching an object with sound*, an object approaches the infant slowly and then moves away quickly. The approaching movement is combined with different vocalizations, showing how one and the same thing – the slow approaching movement – can also be different. In contrast, the withdrawal movement is faster and silent. These small variations and contrasts capture and sustain infant attention. The *Seesaw* reiterated performance matches three movements (leaning back, suspense, and bringing back up) with three sounds. Like in *Smelling paw*, where the baby extends his foot to await the performance of the adult, through the repetition of this routine, the baby gets involved by throwing his head back, summoning the adult to start the game.

The *Barambuch* and *Seesaw* IDI-performances share an active infant participation, who manages to initiate them and request the adult to resume them. In this way, IDI-performances seem to foster the birth of communication intention.

The cases discussed allow us to think that the interplay between the different expressive modalities of early musicality is indifferent to form and combination when it comes to fleeting or reiterated IDI-performances, whether dyadic or familiar. They all share the compositional mode, a simple "strategy for reciprocity," as Español (2004) calls it, of motif repetition-variation.

Protoconversations

Example A – Age of infant: 0; 4 (28). Duration: 26 seconds.

The mother is holding the baby in her arms and they are looking at each other. The mother asks in a very low voice: "Do you have a nanny in your tummy, in your tummy, in your tummy?" The baby looks at her until he responds with a smile to the third maternal vocalization. Then the mother vocalizes louder, longer and higher pitched, smiling very broadly and rocking the child: "Ahhh, who's the pretty one here ... Who's the pretty one?" When the baby does not respond, the mother continues, "Who's the pretty one? ... Who's the pretty one here? ... Who is it, huh?" (maternal turn 1), until the baby finally utters his first vocalization "Ehhh" (infant turn 1). Then she vocalizes again with a marked facial expression, "Ahhh, there you are" (maternal turn 2). The baby, who never stopped looking

at her, vocalizes once more (infant turn 2). After a short pause, the mother vocalizes again looking for a response, making six interventions "Here you are? ... Here you are, baby ... Are you the pretty one here?... Ahhh ... Ahhh, it's you ... It's you." The baby answers with a single sound: "Ahhh," and she smiles back, pleased.

Example B – Age of infant: 0; 3 (23). Duration 30 seconds.

Mother and baby are in the family room. They are face to face, he is lying on top of her legs. The baby looks toward a side wall and vocalizes. The mother takes his hands, pulls his face close, and seeks his attention by saying "Look at me." The baby turns his head sharply to the opposite side and vocalizes, in a whimpering tone, a long "Ahhhhh." The mother moves her head closer to that side, again looking for his gaze and saying, "What? Tell me!" Then the baby looks straight ahead and fixes his gaze on her. "What?" the mother insists, now in a higher pitched tone and with greater intensity, opening her eyes and mouth exaggeratedly. The baby makes whimpering sounds and stretches his body. The mother asks "What ... the belly, the tummy?" as she holds his ankles and begins to shake them gently (maternal turn 1). The baby smiles and vocalizes with intensity, "Ahhhhhh!" (infant turn 1). The mother responds "Ahhhhh!" imitating his utterance and smiling (maternal turn 2). The baby vocalizes again (infant turn 2) and she replies (maternal turn 3). The protoconversation continues for another two cycles (infant turns 3 and 4, maternal turns 4 and 5).

In protoconversations, adults address babies by exaggerating their movements and vocalizations and by making pauses that act as an invitation to participate. In example A, the adult insists on getting the baby involved through the repetition of vocalizations (implying a rhythm and therefore a musical trait), which is complemented with facial expressions. The protoconversation thus demonstrates that its underlying structure responds to the potentialities of musicality and multimodality. In example B, the interaction becomes a protoconversation only when the mother comes into contact with the body of the baby and shakes his lower limbs, activating a stimulation that encompasses movement, vocalization, and facial expression.

In these two examples, we can observe – just as before in the different IDI-performances and as we will observe in action songs (Sect. 3.2.3) – that the emergent organizations of early musicality share the same tools: the scansion of time into rhythmic and reiterative units, and the resource for doing so by virtue of baby multimodal perception. There is a rhythmicity in the vocalizations or movements of each participant, mostly in those of the adult partner (who sometimes repeats her vocalizations or actions as a form of encouragement), but also in the back-and-forth patterns implied in the idea of protoconversation, a rhythmic form where the baby joins in, understanding that a dual activity structure is offered to him, to which he surrenders by entering into performance and pause turns.

Action Songs

It is evident that action songs have a clear multimodality imperative. On the one hand, singing involves the oral modality, either to intone verses engaging language, or for humming. On the other hand, the action corresponds to movements fundamentally oriented to perform what is described in the lyrics (iconic performance) or to mark the beats (pulses, rhythm) of the melody. Between body use and musical characteristics of the piece in question, the emergent organization of action songs

belongs to the world of music and operates simultaneously with several modality options. We will provide a description of each action song (in order of appearance) as we did with the previous emergent organizations.

Cute little hand. The lyrics vary slightly in different Spanish-speaking countries and read approximately: *What a nice little hand I have/what a nice little hand mommy gave me/What a nice little hand I have/I close it, open it and close it again*. The simultaneous action implies starting with a hand extension and simple wrist rotation movements to both sides, then opening and closing, then hiding behind the back, showing it again and hiding it again, all as successively reported in the verses.

Little gray horse. The baby must be seated on the legs of the adult so that he or she can imitate the walking movement of a horse. The song says: *Juan went to Paris/on a little gray horse/in steps, in steps* [slow movement]/*trotting, trotting* [accelerated movement]/*galloping, galloping!* [fast movement]. Eventually the game comes to this ending: *The little horse got tired/and Juan fell off...!*. In this case, the adult abruptly tilts the baby backward, mimicking a fall.

Pimpón the doll. The song narrates everyday actions performed by a doll (*Pimpón is a very handsome cardboard doll/he washes his hands with soap and water [...]/he untangles his hair with an ivory comb/*[...] *Pimpón eats his soup and keeps his apron clean/[...] Pimpón goes to bed and lies down to rest*, etc.). The actions are illustrated through representational gestures. At a certain point, the song asks Pimpón for a handshake (*Pimpón, give me your hand with a strong handshake/I want to be your friend, Pimpón, Pimpón, Pimpón*); in this stanza, the adult takes the hand of the baby, engaging him in a two-way contact movement.

Conga (a kind of dance). The mother sings a conga song while the baby dances flexing and extending his knees successively, bouncing on place. At times the mother holds the hands to support the baby; at other times the baby holds on to the surrounding furniture on his own. The song refrains: *Conga, conga, conga/let the milonga go on/*[pause] *now let's see/how [name of the child] dances the conga*. The mother sings while the child bounces with his knees. If he pauses the movement, she pauses the singing. At the same time, the baby bounces while looking at his mother as if he is trying to get her to sing the song. This is a game song where participants take turns. It is drawn from extra familial culture and transformed into a dyadic expectant interaction. The baby learns that whenever the mother sings the *Conga*, it is his turn to do the bouncing-dance creating a complicity between both parties.

Butter pancakes. The mother claps her hands against each other as she sings*: Pancakes, pancakes/butter pancakes/mama gives me titty/papa gives me cookie* (repeats). The mother claps her hands, keeping the rhythm of the song and waiting for the baby to follow. The baby sees her and participates by joining hands and trying to respect the rhythm she sets. On some occasions, he is the one who starts clapping his hands, looking at the mother so that she starts singing and clapping.

Where is it?. The mother is in the kitchen holding the baby up, looking out the window toward the courtyard. The baby is in her arms in a sitting position; she waves her hands in a joyful gesture and accidentally shakes a curtain next to her with her hand. Then the mother grabs the curtain cloth and covers the baby with it, saying "Where is he?" She then uncovers the child and exclaims, "Here he is!" This action sequence, each one coupling an oral utterance with the appearance and disappearance of the child, has sometimes been registered as hiding objects instead of the child. This is a peek-a-boo variant, described in Chap. 2 as ritualized performance. As we mentioned in the code, this type of interaction was included in the action song category because of its conventional and regulated features, which oppose the predominantly improvised character of the performances.

Eskimo kiss and *Tachín.* These are also cases of ritualized performances involving regulated combinations of sound and movement derived from cultural games. In the first one,

the adult proposes the baby to rub noses: he brings his nose close to the baby and starts rubbing it with his nose from one side to the other, several times, while vocalizing "Ahhhh." In the second, some objects (or the floor, or the body) are hit with the hands while repeating with each stroke, as if the voice would amplify the noise, "Tachín, tachín, tachín." In these types of action songs, much like in the previous ones, there is an absence of ritualized mimicry iconically illustrating the lyrics of the song; nevertheless, they both have a share of pre-established normativity justifying their inclusion in this category.

In these action songs, melodies, words, and action schemes are combined in a narrative sequence identically repeated by the adult in front of the infant. The narrative sequences contain simple actions that the infant gradually incorporates through imitation. The adult encourages the baby to assimilate the action patterns (for example, in *Cute little hand*, moving one hand side to side, opening and closing it, and finally hiding it behind the back) and perform them at the right moment in the narrative, according to the verses. The baby is invited to take part in an arbitrary and conventional way that only makes sense within the limits of that particular action song. As indicated by Eckerdal and Merker (2009), action songs might possibly be a means to introduce the infant to his or her first ritual practice. It may happen that some segment in the narrative sequence cannot be modeled independently by the adult partner but requires some input from the baby. For example, in *Little gray horse*, the final segment of pulling back can only occur if the baby collaborates in performing the movement at the right time. In these cases, the consummation of the routine requires the baby to settle into doing together what the conventional form has placed on his or her side. Emotional communion between the co-protagonists is not achieved by playing in different ways, varying or improvising, but by carrying out the narrative sequence according to a pre-established form. The key is to do it well, which is equivalent to doing it together, in tandem. According to Shifres and Español (2004) during musical play, the adult–infant dyad adjusts its timing to a third member, the metrical structure of songs and/or rhymes. Musical play is coherent with secondary intersubjectivity, where the dyad becomes part of a triangular communion (subject–subject–object). The metrical structure might be considered an abstract object. In a similar way, the action song, with its series of movements and words sung according to a pre-established pattern, is consistent with the mastery of secondary intersubjectivity, where the dyad participates in a triangular communion.

3.2.3 Final Remarks

This research originated in three questions about early social cognition: How do we, as adults, play with our little ones? How do we manage to understand each other, share intimacy and such pleasurable moments of emotional communion? What is so peculiar about communicative musicality behaviors, both nonverbal and verbal, by which we interact with babies and that they seem to enjoy to such an extent? Developmental psychology has been working around these questions since the past century. On the one hand, embodied cognition proposes a mind–body continuum

that constitutes a crucial factor in shaping complex psychological processes. On the other hand, communicative musicality assumes that in early life the melodic contour of speech, the typical timbre variations of the voice, the rhythm or energy of movement, and the scansion of time into repetition and variation units are the basic tools that structure interactions and establish the first adult–infant psychological contacts. Finally, the second-person perspective places intersubjective contact at the center of human development. These three research lines allowed us to identify and characterize the emergent organizations of early communicative musicality, which integrate, within general baby–adult interactions, a set of social phenomena defined by their playful quality, their sound-kinetic composition, and their origin and transformation within ontogenetic development. IDI-performances, protoconversations, and action songs have been studied here in an integrated way and, for the first time to our knowledge, in an empirical way, spanning the period from primary intersubjectivity to the beginnings of secondary intersubjectivity (between 2 and 9 months).

The quantitative data obtained in our single case study show that the three emergent organizations of interactive musicality are present during the first year of life but exhibit different frequencies and developmental trajectories. One of the most striking findings is that IDI-performances were the most frequent phenomenon, followed by action songs and then protoconversations. This information prompts further exploration of IDI-performances while repositioning protoconversation, probably one of the most studied interactive infant phenomena, alongside IDI-performances.

IDI-performances and protoconversations were observed from the beginning of our study (second month of life) as privileged modes of intersubjective contact. From the age of 6 months, however, they followed different trajectories: while IDI-performances reached their peak frequency, protoconversations disappeared. Less frequently in later months, IDI-performances remained throughout the entire studied period as the only one among the emerging organizations being present from the beginning to the end, from 2 to 9 months. Action songs appeared two months after the remaining emerging organizations (at the end of the fourth month), consolidating and achieving their highest frequency by the eighth month.

The period of primary intersubjectivity begins shortly after birth and reaches its peak at around the age of 6 months. Protoconversations are a privileged form of dyadic contact during this period. By 9 months, secondary intersubjectivity – that is, adult–infant contact involving a shared interest in and action on events in the world – is fully established (Trevarthen, 1998). At around 6 months, infants begin to follow the gaze of others and to enjoy playing activities where objects are animated by the mother (Hubley & Trevarthen, 1979). In this way, the dyad transitions into secondary intersubjectivity (see Chap. 5). Action songs emerge during the period of primary intersubjectivity but take hold during the passage to secondary intersubjectivity. Unlike the dominantly unidirectional unfolding of performances and the back and forth of protoconversations, the format of action songs requires that the infant performs certain specific movements and sounds and that he/she does so according to a predetermined and fixed cultural pattern. The comparatively late appearance of action songs could be due to the fact that they require the baby to fit in a particular

way to a predetermined narrative sequence, which may demand more sophisticated cognitive-behavioral skills than merely intervening in the gaps left by the adult partner like in protoconversations, or participating receptively like in IDI-performances.

Action songs must comply with rules that go beyond the dyadic patterns of reciprocity, since they seem unable to do so in the space defined only by the I–You of the dyad. In this sense, they tend to function as a third element, an external object, to which the dyad attends, shares, and abides. The action songs – initially unfolded by the adult to which the infant gradually joins in by imitating the sound and motor adult behavior – could be aiding the transition toward secondary intersubjectivity.

While protoconversations are privileged modes of primary intersubjectivity contact and action songs are privileged modes of secondary intersubjectivity, performances cut across both modes. As we noted in Chap. 2, IDI-performances seem to serve different functions, from mood regulation to image-schema enhancement, from the intersubjective and intercorporeal experience of primary metaphors to the aesthetic and bodily infant enculturation. It is perhaps because of this multifunctionality that they accompany the development of primary and secondary intersubjectivity.

The interpersonal world of the infant is constituted by the exchanges taking place in milliseconds or a few seconds (Stern, 1985, 2010, and Chap. 4 in this volume). Emergent organizations extend the time of adult–infant exchange units. In our study, none of them exceeded 30 seconds. This data is a clear indicator that even in the most extended exchange moments, the interpersonal world of the infant remains a seconds-long world. While the differences found in the duration of the three organizations were not statistically significant, protoconversations were the longest on average, followed by IDI-performances, and third by action songs. Protoconversations may have required longer periods of time: the adult usually vocalized while looking at the baby, as if waiting for his reaction; he made long pauses, leaving some room for the baby to feel challenged, summoned to respond. On the other hand, performances require diachrony rather than simultaneity (the modalities converge in time and content: they synchronically display gaze, movement, and vocalization behaviors organized in a repetition-variation structure). Although this particular characteristic is not entirely absent in protoconversations, it is more present in performances. In action songs, there is an obligatory diachronic unfolding but governed by a preestablished sound-kinetic script.

While elaborating the observation code, it seemed pertinent to distinguish between fleeting performances (that appear, achieve a good form, and then vanish) and reiterated performances that nurture an interactive repertoire to which the adult turns to on different occasions. Each of these types offers the baby different experiences. Because of its novelty, in the fleeting type of IDI-performance, the baby can anticipate future experiences only in the microcosm of repetition-variation forms. Whereas in the reiterated IDI-performances, this anticipatory exercise is coupled with the concomitant enjoyment of identifying an interactive behavioral pattern already experienced, allowing the infant to participate more actively and request its repetition (as seen in *Seesaw*).

Among the IDI-performances, reiterated IDI-performances were the most frequent. The preference of the adult for these routines could be due to a possible detection in the infant of a certain recognition of previously practiced forms. The presumed sense of pleasure caused by the identification could unconsciously appeal to the inclination of the adult toward repeated routines. This could be the reason for the frequency shift from fleeting performances, dominant in the first sessions, to repeated ones, which gradually consolidate their presence. We must always consider that adults are responsive to the reactions of infants and that the satisfaction of babies derived from the interaction, particularly when they identify a previously known multimodal play structure, would encourage the adult to repeat these performances. Noticing that the infant recognizes a formerly executed performance could be functioning as a motivation to gradually replace fleeting with repeated ones.

The sound-kinetic scripts of reiterated performances were simpler than those of action songs. In dyadic or familiar reiterated performances, the actions and movements of the adult usually involved one single element. For example, in *Barambuch* the action would only consist in the adult slowly bringing his head close to the head of the baby, stressing the movement with vocalizations until contact was established. In contrast, the singing and movement script in the action song required a chain of different components operating synchronously. *Little gray horse*, *Cute little hand*, and *Pimpón the doll* contain sequences of three, four, or more units of meaning.

In general terms, fleeting performances showed that the combination of musical features of sound-kinetic units (ascent and descent in the scale of notes, accelerations and decelerations, alternation of segments in *fortissimo* and *pianissimo*, etc.) is indistinctly transferred to the sonorous or kinetic modalities or to both. When a feature, for example, the pulse, falls in the oral sonorous modality (*Co...go...te* performance), the motricity assumes the mission of integrating the three beats of the vocalization into one and the same motif (similar to what happens in the familiar *Stinguishen* reiterated performance, where the vocalization fits a number of times in the longitudinal body trajectory carried out by the hands of the adult). Conversely, like in example B of fleeting performances (hand to mouth), movements can also be triggered and interrupted with a unitary, extended vocalization ("Baaaaa..."). Or vocalization and movement might perfectly coincide with each other, rounding off an independent musical motif (example C of fleeting performances). We were also able to register how the back and forth or reversibility of a movement was highlighted by two beats of the same vocalization ("Tu-tu" in example C). These three forms of intermodal integration were observed in both reiterated performances, dyadic and familiar, including new elements such as the use of silence to emphasize movement opposition (*Gaze interplay*). The movement acceleration and explosive vocalization-movement finale were observed in different routines (more clearly in *Barambuch*).

Usually, the IDI-performance is characterized by adult-directed infant attention and the execution of social behaviors (such as smiling, gurgling, vocalizations) which makes him/her a co-participant in the exchange. He/she is neither an active participant nor a mere observer. The adult's multimodal-behavioral display may

sometimes contain kissing and tickling, rocking, holding, and moving in different ways rather than being just a distant action where infants are simply spectators. Audiovisual multimodality is nourished by proprioception, a different source of information. IDI-performances involve interactions that allow babies to contemplate or to experiment with their whole body. Indeed, our research featured examples involving an active engagement on his part, for example, in *Smelling paw* (where he extended his foot for the adult to smell) and *Seesaw* (where he threw his head back eliciting the play). Protoconversations, on the other hand, only exist if the baby unfailingly takes an active participation. Strictly speaking, it is the back-and-forth structure (alternation of turns) that characterizes this form of interaction. On the other hand, the structure of action songs allows for an involvement oscillating between two extremes (passive observer or active role), even when it seems to be generally designed to promote a co-participative performance through imitation. To our knowledge, none of the three types of emergent organizations (protoconversation, performance, and action songs) has been described in terms of their developmental variability in the first months of baby life.

With single-case studies, we can only aspire to draw non-generalizable conclusions, subject to validation or refutation by subsequent work. IDI-performances are still a terrain of incipient exploration and, for its part, the idea of emergent organizations as a particular set of interactions in infancy and early childhood must be subjected to future empirical studies. Perhaps further research will isolate other variants of emergent organizations still undetected. We have sought to share our approach to these particular forms of interaction in the interpersonal world of infants.

References

Bainbridge, C. M., Bertolo, M., Youngers, J., Atwood, S., Yurdum, L., Simson, J., Lopez, K., Feng, X., Martin, A., & Mehr, S. A. (2021). Infants relax in response to unfamiliar foreign lullabies. *Nature Human Behaviour, 5*(2), 256–264.

Bateson, M. C. (1971). The interpersonal context of infant vocalization. *Quarterly Progress Report of the Research Laboratory of Electronics, 100*, 6–170.

Bateson, M. C. (1975). Mother-infant exchanges: The epigenesis of conversational interaction. *Annals of the New York Academy of Sciences, 263*(1), 101–113. https://doi.org/10.1111/j1749-6632.1975.tb41575.x

Bateson, M. C. (1979). The epigenesis of conversational interaction: A personal account of research development. En M. Bullowa (Ed.), Before speech. The beginning of interpersonal communication (pp. 63–78). : Cambridge University Press.

Beebe, B. (1982). Micro-timing in mother-infant communication. In M. R. Key (Ed.), *Nonverbal-communication today* (pp. 169–195). Mouton Publishers.

Beebe, B. (2014). My journey in infant research and psychoanalysis: Microanalysis, a social microscope. *Psychoanalytic Psychology, 31*(1), 4–12. https://doi.org/10.1037/a0035575

Beebe, B., & Gerstman, L. J. (1980). The "packaging" of maternal stimulation in relation to infant facial-visual engagement: A case study at four months. *Merrill-Palmer Quarterly of Behavior and Development, 26*(4), 321–339.

Beebe, B., Jaffe, J., Feldstein, S., Mays, K., & Alson, D. (1985). Interpersonal timing: The application of an adult dialogue model to mother-infant vocal and kinesic interactions. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 217–247). Albex.

Beebe, S., Lachmann, F. M., Markese, S., Buck, K. A., Bahrick, L. E., Chen, H., Cohen, P., Andrew, H., & Jaffe, J. (2012). On the Origins of Disorganized Attachment and Internal Working Models: Paper II. An Empirical Microanalysis of 4-Month Mother–Infant Interaction. *Psychoanalytic Dialogues: The International Journal of Relational Perspectives*, 352–374. https://doi.org/10.1080/10481885.2012.67960

Bloom, K. (1977). Patterning of infant vocal behavior. *Journal of Experimental Child Psychology, 23*(3), 367–377. https://doi.org/10.1016/0022-0965(77)90032-7

Boiteau, C., Kokkinaki, T., Sankey, C., Buil, A., Gratier, M., & Devouche, E. (2021). Father–newborn vocal interaction: A contribution to the theory of innate intersubjectivity. *Infant and Child Development, 30*(5), e2259.

Brand, R., & Shallcross, W. (2008). Infants prefer motionese to adult-directed action. *Developmental Science, 11*(6), 853–861.

Brand, R., Baldwin, D., & Ashburn, L. (2002). Evidence for 'motionese': Modifications in mothers' IDI-performance action. *Developmental Science, 5*(1), 72–83.

Bråten, S. (2009). *The intersubjective mirror in infant learning and evolution of speech*. John Benjamins Publishing.

Bullowa, M. (1979). Introduction: Prelingüistic communications: A field for scientific research. En M. Bullowa. *Before speech. The beginning of interpersonal communication* (pp. 8–16). : Cambridge University Press.

Calmels, D. (2014). *El cuerpo cuenta. La presencia del cuerpo en las versificaciones, narrativas y lecturas de crianza*. Homo Sapiens.

Calvo, P., & Gomila, T. (Eds.). (2008). *Handbook of cognitive science: An embodied approach*. Elsevier.

Condon, W. S., & Sander, L. W. (1974). Synchrony demonstrated between movements of the neonate and adult speech. *Child Development, 45*(2), 456–462. https://doi.org/10.2307/1127968

Corbeil, M., Trehub, S. E., & Peretz, I. (2013). Speech vs. singing: Infants choose happier sounds. *Frontiers in Psychology, 4*. https://doi.org/10.3389/fpsyg.2013.00372

Corbeil, M., Trehub, S. E., & Peretz, I. (2016). Singing delays the onset of infant distress. *Infancy, 21*(3), 373–391.

Delack, L. B., & Fowlow, P. J. (1978). The ontogenesis of different vocalizations: Development of prosodic contrastivity during the first year of life. In N. Waterson & C. Snow (Eds.), *The development of communication* (pp. 93–110). Wiley.

Demuth, C. (2013). Protoconversation and protosong as infant's socialization environment. In T. M. S. Tchombe, B. Nsamenang, H. Keller, y M. Fülöp (Eds.), Cross-cultural psychology: An Africentric perspective (pp. 232–256). : Design House Ed.

Devouche, E., & Gratier, M. (2019). The beginning of parent-infant communication. In G. Apter, E. Devouche, & M. Gratier (Eds.), *Early interaction and developmental psychopathology* (pp. 21–33). Springer. https://doi.org/10.1007/978-3-030-04769-6_2

Dissanayake, E. (2000a). *Art and intimacy: How the arts began*. Washington University Press.

Dissanayake, E. (2000b). Antecedents of the temporal arts in early mother-infant interaction. In N. L. Wallin, B. Merker, & S. Brown (Eds.), *The origins of music*. The MIT Press.

Dominguez, S., Devouche, E., Apter, G., & Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant and Child Development, 25*(3), 240–255. https://doi.org/10.1002/icd.1976

Durt, C., Fuchs, T., & Tewes, C. (Eds.). (2017). Embodiment, enaction, and culture: Investigating the constitution of the shared world. MIT Press.

Eckerdal, P. & Merker, B. (2009). "Music" and the "action son" in infant development: An interpretation. En S. Malloch & C. Trevarthen (Eds.), Communicative musicality: Exploring the basis of human companionship (pp. 241–258). : Oxford University Press.

Edward, Tronick Heidelise, Als Lauren, Adamson Susan, Wise T. Berry, Brazelton (1978) The Infant's Response to Entrapment between Contradictory Messages in Face-to-Face

Interaction. Journal of the American Academy of Child Psychiatry 17(1) 1-13 10.1016/S0002-7138(09)62273-1

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana* [How to do things without words. Gesture and fiction in early infancy]. Antonio Machado.

Español, S., & Shifres, F. (2015). The artistic infant directed performances: A mycroanalysis of the adult's movements and sounds. *Integrative Psychological and Behavioral Science, 49*(3), 371–397. https://doi.org/10.1007/s12124-015-9308-4

Español, S., Bordoni, M., Martínez, M., Camarasa, R., & Carretero, S. (2015). Forms of vitality play and symbolic play during the third year of life. *Infant Behavior and Development, 40*, 242–251. https://doi.org/10.1016/j.infbeh.2015.05.008

Fernald, A. (1989). Intonation and communicative intent in mothers' speech to infants: Is the melody the message? *Child Development, 60*, 1497–5100.

Fogel, A., & Garvey, A. (2007). Alive communication. *Infant Behavior and Development, 30*, 251–257. https://doi.org/10.1016/j.infbeh.2007.02.007

Fogel, A., Toda, S., & Kawai, M. (1988). Mother-infant face-to-face interaction in Japan and the United States: A laboratory comparison using 3-month-old infants. *Developmental Psychology, 24*(3), 398.

Forster, S., & Iacono, T. (2013). The nature of affect attunement used by disability support workers interacting with adults with profound intellectual and multiple disabilities. *Journal of Intellectual Disability Research, 58*(12), 1105–1120.

Garvey, A., & Fogel, A. (2007). Dialogical change processes, emotions, and the early emergence of self. En H. J. M. Hermans (Ed.). *International Journal for Dialogical Science, 2*, 51–76.

Ginsburg, G. P., & Kilbourne, B. K. (1988). Emergence of vocal alternation in mother-infant interchanges. *Journal of Child Language, 15*(2), 221–235.

Goldberg, S. (1977). Social competence in infancy: A model of parent-infant interaction. *Merrill-Palmer Quarterly of Behavior and Development*, 163–177.

Gratier, M., & Apter-Danon, G. (2009). The improvised musicality of belonging: Repetition and variations in mother-infant vocal interaction. En S. Malloch, & C. Trevarthen (Eds.), Communicative musicality: Exploring the basis of human companionship, (pp. 281–300). : Oxford University Press.

Gratier, M., & Trevarthen, C. (2008). Musical narrative and motives for culture in mother-infant vocal interaction. *Journal of Consciousness Studies, 15*(10), 122.

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology, 6*(1167), 10–3389. https://doi.org/10.3389/fpsyg.2015.01167

Grice, P. (1989). *Studies in the way of words*. Harvard University Press.

Glaser, B. & Strauss, A. (1967). The Discovery of Grounded Theory. New York: Aldine Publishing Company.

Gogate, L. J., Bahrick, L. E., & Watson, J. D. (2000). A study of multimodal motherese: The role of temporal synchrony between verbal labels and gestures. Child development, 71(4), 878–894.

Hernandez-Reif, M., Diego, M., & Field, T. (2007). Preterm infants show reduced stress behaviors and activity after 5 days of massage therapy. *Infant Behavior and Development, 30*(4), 557–561.

Hilbrink, E. E., Gattis, M., & Levinson, S. C. (2015). Early developmental changes in the timing of turn-taking: A longitudinal study of mother–infant interaction. *Frontiers in Psychology, 6*, 1492.

Hsu, H. C., & Fogel, A. (2003). Stability and transitions in mother-infant face-to-face communication during the first 6 months: A microhistorical approach. *Developmental Psychology, 39*(6), 1061.

Hubley, P. Trevarthen C. (1979). Sharing a task in infancy. En I. Uzgiris (Ed.), Social interaction during infancy: New directions for child development, vol. 4 (pp. 57–80). : Jossey Bass.

Imberty, M. (2002). La musica e il bambino. In J. J. Nattiez (Ed.), *Enciclopedia della Musica* (pp. 477–495). Giulio Einaudi Editore.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Developmental*, i–149.

Johnson, M. (2007). *The meaning of the body. Aesthetics of human understanding*. The University of Chicago Press.

Jonsson, C. O., Clinton, D., Fahrman, M., Mazzaglia, G., Novak, S., & Sörhus, K. (2001). How do mothers signal shared feeling-states to their infants? An investigation of affect attunement and imitation during the first year of life. *Scandinavian Journal of Psychology, 42*(4), 377–381.

Kaye, K. (1982). *The mental and social life of babies: How parents create persons*. Methuen.

Kiepura, E., Niedźwiecka, A., & Kmita, G. (2022). Silence matters: The role of pauses during dyadic maternal and paternal vocal interactions with preterm and full-term infants. *Journal of Child Language, 49*(3), 451–468. https://doi.org/10.1017/S030500092100012X

Kokkinaki, T. (2010). Intersubjectivity during free infant-father 'protoconversation' and within-'protoconversation' pauses. *Early Child Development and Care, 180*, 1–2. https://doi.org/10.1093/deafed/5.2.127

Lavelli, M., & Fogel, A. (2013). Interdyad differences in early mother–infant face-to-face communication: Real-time dynamics and developmental pathways. *Developmental Psychology, 49*(12), 2257.

Leman, M. (2008). *Embodied music cognition and mediation technology*. MIT Press.

LeVine, R. A., Dixon, S., LeVine, S., Richman, A., Keefer, C. H., & Brazelton, T. B. (1996). *Child care and culture: Lessons from Africa*. Cambridge University Press.

Lyra, M. C., & Valsiner, J. (2011). Historicity in development: Abbreviation in mother-infant communication. *Infancia y Aprendizaje, 34*(2), 195–203. https://doi.org/10.1174/021037011795377638

Lyra, M. C. D. P. (2007). On abbreviation: Dialogue in early life. International Journal for Dialogical Science, 2(1), 15–44.

Malloch, S., & Trevarthen, C. (Eds.). (2009). *Communicative musicality: Exploring the basis of human companionship*. Oxford University Press.

Martínez, I. (2014). La base corporeizada del significado musical. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* (pp. 71–110). Paidós.

Martínez, I., Español, S., & Pérez, D. I. (2018). The interactive origin and the aesthetic modelling of image-schemas and primary metaphors. *Integrative Psychological and Behavioral Science, 52*, 646–671. https://doi.org/10.1007/s12124-018-9432-z

Mehr, S. A., Singh, M., Knox, D., Ketter, D. M., Pickens-Jones, D., Atwood, S., … Glowacki, L. (2019). Universality and diversity in human song. *Science, 366*(6468), eaax0868.

Murray, L., & Trevarthen, C. (1985). Emotional regulation of interactions. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 177–197). Ablex.

Murray, L., & Trevarthen, C. (1986). The infant's role in mother–infant communications. *Journal of Child Language, 13*(1), 15–29.

Oller, D. K., Caskey, M., Yoo, H., Bene, E. R., Jhang, Y., Lee, C. C., & Vohr, B. (2019). Preterm and full term infant vocalization and the origin of language. *Scientific Reports, 9*(1), 1–10.

Ospina Tascón, V., & Español, S. (2014). El movimiento y el sí mismo. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* (pp. 111–155). Paidós.

Papoušek, H. (1996a). Musicality in infancy research: Biological and cultural origins of early musicality. En I. Deliège, J. A. Sloboda (Eds.), Musical beginnings (pp. 37–55). : Oxford University Press.

Papoušek, M. (1996b). Intuitive parenting: A hidden source of musical stimulation in infancy. En I. Deliège & J. Sloboda (Eds.). Musical beginnings. Origins and development of musical competence. (pp. 88–112). : Oxford University Press.

Papoušek, M., & Papoušek, H. (2002). Intuitive parenting. In M. H. Bornstein (Ed.), *Handbook of parenting. Volumen 2. Biology and ecology of parenting* (pp. 183–203). Lawrence Erlbaum.

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Rivière, A., & Sotillo, M. (1999/2003). Comunicación, suspensión y semiosis humana: los orígenes de la práctica y de la comprensión interpersonal En: M. Belinchón, A. Rosa, M. Sotillo e I. Marichalar (comp.). In *Ángel Rivière. Obras Escogidas, Vol III* (pp. 181–201). Panamericana.

Rochat, P., Querido, J. G., & Striano, T. (1999). Emerging sensitivity to the timing and structure of protoconversation in early infancy. *Developmental Psychology, 35*(4), 950. https://doi.org/10.1037/0012-1649.35.4.950

Rowlands, M. (2010). *The new science of the mind*. From Extended Mind to Embodied Phenomenology. Bradford. MIT Press.

Savage, P. E., Loui, P., Tarr, B., Schachner, A., Glowacki, L., Mithen, S., & Fitch, W. T. (2021). Music as a coevolved system for social bonding. *Behavioral and Brain Sciences, 20*, 44–59.

Schieffelin, B., & Ochs, E. (1995). Language acquisition and socialization: Three developmental stories and their implications. In R. Shweder & Levine (Eds.), *Essays on mind, self and emotion* (pp. 276–230). Cambridge University Press.

Sheets-Johnstone, M. (2011). *The primacy of movement* (Vol. 82). John Benjamins Publishing.

Sheets-Johnstone, M. (2015). *The corporeal turn: An interdisciplinary reader*. Andrews UK Limited.

Shifres, F. (2011). La expresión en el canto dirigido al bebé. *Psicología del Desarrollo, 1*(2), 29–42.

Shifres, F. & Español, S. (2004). Interplay between pretend and music play. *8th International Conference on Music Perception & Cognition. European Society for the Cognitive Science of Music* (ESCOM), Evanston.

Snow, C. E., & Ferguson, C. A. (1977). *Talking to children: Language input and acquisition*. Cambridge University Press.

Stern, D. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. En M. Lewis, & L.A. Rosenblum (Eds.), The effect of the infant on its caregiver (pp. 402–421). : Wiley-Interscience.

Stern, D. (1985). *The interpersonal world of the infant. A view from psychoanalysis and developmental psychology*. Basic Books.

Stern, D. (2010). Forms of vitality. In *Exploring dynamic experience in psychology, arts, psychotherapy and development*. Oxford University Press.

Stern, D., Jaffe, J., Beebe, B., & Bennet, S. (1975). Vocalizing in unison and alternation. *Transactions: New York Academy of Sciences, 263*, 89–101. https://doi.org/10.1111/j.1749-6632.1975.tb41574.x

Strauss, A., & Corbin, J. (1994). Charpter 17: Grounded theory methodology: An overview. In N. K. Denzin & Y. S. Lincoln (Eds.), *Handbook of qualitative research* (pp. 273–285). SAGE.

Soledad Carretero, Pérez Silvia Ana, Español (2016) Multimodal Study of Adult-Infant Interaction: A Review of Its Origins and Its Current Status. Paidéia (Ribeirão Preto) 26(65) 377–385 10.1590/1982-43272665201613

Stern, D. & Gibbon, J. (1979). Temporal expectancy during mother-infant play. In Thoman, E.. (ed.), Origins of the infant's social responses (pp. 409-429). New York: Erlbaum.

Striano, T., Henning, A., & Stahl, D. (2005). Sensitivity to social contingencies between 1 and 3 months of age. Developmental Science, 8(6), 509–518.

Trehub, S. E., Unyk, A. M., Kamenetsky, S. B., Hill, D. S., Trainor, L. J., Henderson, J. L., & Saraza, M. (1997). Mothers' and fathers' singing to infants. *Developmental Psychology, 33*(3), 500.

Trehub, S. E., Plantinga, J., Brcic, J., & Nowicki, M. (2013). Cross-modal signatures in maternal speech and singing. *Frontiers in Psychology, 4*(November), 811. https://doi.org/10.3389/fpsyg.2013.00811

Trehub, S. E., & Gudmundsdottir, H. R. (2015). *Mothers as singing mentors for infants. Online publication*. Oxford University Press.

Trehub, S. E., & Cirelli, L. K. (2018). Precursors to the performing arts in infancy and early childhood. *Progress in Brain Research, 237*, 225–242.

Trevarthen, C. (1977). Descriptive analysis of infant communicative behaviour. En H. R. Schaffer (Ed.), Studies in mother-infant interaction: Proceedings of the Loch Lomond symposium (pp. 227–270). : Academic Press.

Trevarthen, C. (1979). Communication and cooperation in early infancy. En M. Bulowa (ed.) Before speech. The beginning of interpersonal communications (pp. 321–347). : Cambridge University Press.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. En S. Braten (Ed.), *Intersubjective communication and emotion in early ontogeny*. : Cambridge University Press.

Trevarthen, C. (2000). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicae Scientiae, 3*(1 suppl), 155–215. https://doi.org/10.1177/10298649000030S109

Trevarthen, C., & Reddy, V. (2007). Consciousness in infants. En M. Velmans & S. Schneider (Eds.), The Blackwell companion to consciousness (pp. 41–57). Blackwell Publishing.

Trevarthen, C., & Schögler, B. (2007). Dancing minds. In S. Braten (Ed.), *Being moved by action, perception, music and speech. Mirror neurons and the developmental bases of (pre)verbal intersubjectivity and clinical applications*. Cambridge University Press.

Weiss, M. W., Trehub, S. E., & Schellenberg, E. G. (2012). Something in the way she sings: Enhanced memory for vocal melodies. *Psychological Science, 23*, 1074–1078. https://doi.org/10.1177/0956797612442552

Soledad Carretero has a PhD. in Psychology from Universidad Nacional de Córdoba (UNC, Argentina). She is a Faculty member (FLACSO, Argentina, and Universidad Europea, Madrid, Spain). She has a Degree in Improvised Dancing (Free Dance) from Universidad Nacional del Arte (Argentina). She teaches in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO). She teaches in FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. Her work deals with cognitive developmental psychology and the study of the communicative and therapeutic power of human companionship through movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/soledad.carretero.perez

Silvia Español holds a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). She is currently a researcher at the Latin America Institute of Social Research (IICSAL), Faculty of Latin America Social Sciences/ National Scientific and Technical Research Council (FLACSO/CONICET, Argentina). She coordinates the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc. in Cognitive Psychology, and Director of FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is a certified Somatic Educator in the Feldenkrais Method. Her work deals with cognitive developmental psychology, psychology of music, and the study of human movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/silvia.espanol

Fernando G. Rodríguez is PhD. in Educational Sciences from Universidad de Buenos Aires. He has carried out research and training stays in Germany (DAAD) in Linguistics and Semiotics. He is currently Professor of Semiology and Human Communication. He is member of the Research Group of Socio-Cognitive Processes in Infancy and Early Childhood at the Psychology of Knowledge and Learning Department (FLACSO, Argentina). As a researcher, his work is concerned with the development of communicative skills in young children, the structure of semiotic processes in general and the relationships between mind, sign and corporeality. Author of *Semántica y pragmática en la comunicación del niño pregramatical. Gesto y palabra antes de la articulación morfosintáctica* [Semantics and pragmatics in the communication of the

pregrammatical child. Gesture and speech before morphosyntactic articulation] (2021) and co-editor of *Los signos del cuerpo. Enfoques multimodales de la mente y el lenguaje* [The signs of the body. Multimodal approaches to mind and language] (2018). His publications can be found at https://www.aacademica.org/fernando.gabriel.rodriguez

Favio Shifres is a pianist, chamber music teacher and orchestral conductor from the National University of La Plata (UNLP, Argentina), and PhD. from the University of Roehampton (United Kingdom). He is a regular Professor in Auditory Education and Comparative Music Education (UNLP), and Psychology of Music and Arts in several postgraduate courses (Universidad nacional dde las Artes, UNA; FLACSO; UNLP, Universidad Nacional de Rosario, UNR). Director of research projects. Editor in Chief of Epistemus. Journal of music, cognition and culture. Guest editor in numerous research journals. Founding member and former president of the Argentine Society for the Cognitive Sciences of Music (SACCoM).

Chapter 4
Early Reciprocity: Temporal Coordination and Modality Behavior in Parental Imitation and Affect Attunement

Silvia Español, Mariana Bordoni, Mauricio Martínez, Santiago García Cernaz, and Viviana Riascos

Abstract The interpersonal world assumes a rapid, dynamic, reciprocal, body-present exchange with whom we interact. Adults are often responsive to infant behavior from the first months of life and thus scaffold their learning of interpersonal reciprocity. Certain adult responses involve some kind of matching of the behavior of the infant. Maternal imitation and affect attunement are two recognized matching activities that set a special degree of intersubjectivity and mutuality. In this chapter, we present a single case study of a mother–baby dyad where we analyze both matching activities between the 2nd and 10th month of the life of the infant. Maternal imitation is a phenomenon widely studied in cognitive developmental psychology; affect attunement, on the other hand, despite being a recognized

S. Español (✉)
Latin American Institute of Social Research (IICSAL), Latin American Faculty of Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina
e-mail: sespanol@flacso.org.ar

M. Bordoni
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

M. Martínez
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

S. G. Cernaz
National Scientific and Technical Research Council/ Laboratory for the Study of Musical Experience, Faculty of Fine Arts, National University of La Plata (CONICET/ LEEM-FBA-UNLP), La Plata, Argentina

V. Riascos
Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_4

phenomenon, has been scarcely researched empirically. Our research offers novel empirical data on affect attunement and findings on imitation in dialogue with previous studies. We focus on the development of maternal imitation and affect attunement frequencies, the temporal coordination between infant behavior and adult response, and the pairing between the behavioral modalities of the baby and the mother (vocalization, facial expression, movement).

Keywords Imitation · Affect attunement · Reciprocity · Movement temporal coordination · Maternal responsiveness · Matching activities

4.1 Introduction

The interpersonal world implies a reciprocal, body-present, rapid, and dynamic exchange with those we interact with. We are all familiar with the embarrassment or discomfort arising if our fingers collide when shaking hands with someone; or the uncomfortable, sometimes funny awkwardness when we anticipate a collision with someone who is walking in front of us and we move laterally as the other person does so in the same direction, to find ourselves again in a collision situation. We are all alarmed or uncomfortable when someone holds a *still-face* while interacting with us. In all of these circumstances, the fluid reciprocity of the interpersonal world breaks down. We clearly notice that the mutual accommodation characterizing the reciprocity of body-present interpersonal encounters has failed.

Reciprocity, which passes almost unnoticed when unbroken, underpins our ability of being together. Although we can also have intersubjective experiences beyond body-present encounters (for example, when inferring the beliefs of someone else, when reading a letter from a loved one, or while reading a novel), the embodied, implicit knowledge of body-present encounters precedes and supports the capacities for distant psychological contact.

4.1.1 Reciprocity and Second-Person Perspective

In recent decades, the *second-person perspective in social cognition* emphasized how necessary the more basic and early modes of social engagement are for the development of more distal ones, as well as the lifelong enduring nature of the most basic ones (see Chap. 1 for a description of the second-person perspective in social cognition as an alternative approach to the traditional first- and third-person perspectives). From this paradigm, the basic ways of understanding others have been pointed out as settled on a set of reciprocal understanding skills that develop and are expressed in interactive contexts. Thus, the behavior shown by a subject elicits a sensitive response in others, in a way that actions and reactions share the traits of reciprocity. The expressive aspects of the behavior of either person – such as the

voice, muscle tone, or smile – are perceived as directly significant and constitute the basis for a fully meaningful corresponding reaction within the interaction (Gomila, 2002). In second-person encounters, "(...) what one does elicits a proper response from the other, a response that in turn impacts on the first person, in a mutual, continuous, dynamic, and fluid interplay" (Pérez & Gomila, 2021, p. 16).

This description, fully adequate for most body-present interpersonal exchanges, would not seem to appropriately characterize the earliest body-present encounters in infant life. As noted in Chap. 2, within developmental psychology, reciprocity is widely recognized as the essence of interpersonal encounters, yet adult–baby interaction symmetry remains a matter of debate. Some emphasize that early exchanges are bidirectional yet asymmetrical: caregivers are generally responsible for maintaining dyadic synchrony, whereas infants tend to carry interactions through the induction of contingent responses to their own behaviors by the caregiver (Beebe et al., 2016). Others consider exchanges to be symmetrical from the beginning: adult behavior is part of a developmental strategy centered on the intentional behavior of the infant with whom they co-regulate their participation (Trevarthen, 2017). According to Provenzi et al. (2018), the term *reciprocity* predominates when exchanges are thought to be symmetrical, and the term *mutuality* when they are considered to be bidirectional but asymmetrical.

We believe the first contact experiences – which we like to call mutuality – emerge early in the life of infants through phylogenesis, gestation, and the encounter with adults who receive and care for them. Following Kaye (1982), we also believe that reciprocity – understood as the fluid interactive back-and-forth where our behavior elicits an appropriate response in others in a dynamic exchange loop – is experienced very early in life; however, it is not given to us and must be acquired. To our understanding, from the primary experiences of mutuality, the first six months of life are largely devoted to reciprocity learning. As Schaffer (1977) pointed out, infants will learn that relationships are reciprocal, that dialogues are shaped by both participants, and that their roles are interchangeable (as in a *give and take* game).

At birth, infants have biological rhythms, such as the sucking-pausing rhythm in breastfeeding; they have perceptual preferences or biases toward stimulus parameters that characterize people, such as facial configuration, the sound of language, and musical sounds (Reddy, 2008a, b, Rivière, 1986/2003, Stern, 1985, 2010, Trevarthen, 1998, 1999, Chap. 2 in this volume); they also have movement skills (Stern, 2010); perceptual modes that fit perfectly with adult behavior multimodality, such as multisensory perception (see Chap. 5 and 6); and a fine sensitivity to stimuli contingent upon their own behaviors (Bigelow & Rochat, 2006; Reddy et al., 2007). The baby carries this phylogenetically inherited or in-utero developed baggage (from which we already mentioned some examples), but must organize it in his/her coupling with the world and meet people who can flexibly accommodate his/her rhythms, perceptual biases, movements, vocalizations, to enter together into primary experiences of mutuality (such as breastfeeding or extended mutual gazing). Infants must also meet people who, supported by mutuality experiences, scaffold their reciprocity learning. Multiple initial coupling modes have been described, as well as gradually increasing symmetrical adult–infant exchanges. Among other

things, adults (a) flexibly accommodate the sucking-pausing rhythms of the newborn when nursing (Kaye, 1982), allowing a *primitive nurturing* or *biological conversation* to emerge wherein the learning of reciprocity begins (Español, 2014; Rivière, 1986/2003); (b) summon infants into social life and socially engage them with resources specific to temporal arts (see Chap. 2); (c) modulate the vitality forms of their own movements and vocalizations by regulating the level of infant arousal (see Chaps. 2 and 9); and (d) hold the gaps in turn-taking of early proto-conversations, when infants cannot yet bridge them, thereby enhancing their gradual entry into turn-taking alternation (see Chap. 3). With these and other resources, caregivers lead infants to the *perceptual bonding of active reciprocity* that Reddy (2008b) considers characterizing second-person encounters. The learning of reciprocity is dizzying. Its speed and spontaneity probably relate to the fact that infants, all their senses open to the world, are oriented toward adults who, in addition to being available to interact socially, guide them in a nonconscious and nonvoluntary way, through an intuitive know-how where biology and culture are intertwined (see Chap. 2).

Research on dyad reciprocity or mutuality generally focuses on certain interactive processes like attunement, contingency, coordination, matching, mirroring, reparation, and synchrony (Provenzi et al., 2018). Some of them will be explored in this chapter. Given our assumption on the adult scaffolding of infant reciprocity learning, we are particularly interested in observing adult responsiveness in these processes, i.e., the availability and organization of adult behavior when responding to infant behavior.

4.1.2 Adult Responsiveness

The study of early infant–adult interactions includes the analysis of infant and adult behaviors, the relationship between them, and the context or environment where they occur. Given its complexity, studies sometimes focus on certain elements at the expense of others. Likewise, early infant interactions may be studied in different temporal dimensions. In Chaps. 2 and 3, concentrating on adult behavior, we paid attention to time-extended phenomena – such as *infant-directed improvised performances* or *proto-conversations* – within which shorter, smaller behavioral units exist and are organized. Thus, a brief behavior imitation may enter into mutual imitation cycles, or cycles of other reciprocal behaviors, enabling an extensive proto-conversation; or may lead to a motif whose varied reiteration organizes an infant-directed improvised performance. In other words, within *extended duration formations*, we can recognize shorter units determining the flow of reciprocity; at the same time, we can recognize short units determining the flow of reciprocity beyond temporally extended phenomena. In extended duration formations, reciprocity in the shortest units bears its unfolding; that is to say, the back and forth flow within interactions organizes around the concatenation of small behavioral units (vocalizations, smiles, imitations, movements, etc.) of adults and infants. On other

occasions, the brief reciprocal events simply weave the changing flow of social interaction.

Adult availability for infant social engagement brings a variety of intuitive, non-conscious resources into play. These resources unleash and come to life when we are in the presence of a baby. In this caregiving activity – imperative to the survival and psychological/body wellbeing of infants – adults scaffold infant development and reciprocity learning; and many times, access incredible experiences of wellbeing themselves. In Chaps. 2 and 3, we analyzed the resources through which the nurturing figures create – with and for infants – formations of extensive duration: proto-conversations, infant-directed improvised performances, and action songs. In this chapter, we intend to broaden the study of adult reciprocity resources in short-duration units. This implies focusing on the most basic or atomic units of adult–infant interaction. As indicated by Van Egeren et al. (2001), since the seminal description of moment-to-moment interactions within the mother–infant dyad made by Daniel Stern (1974), researchers are gradually constructing a picture of dynamic collaborations from which predictable patterns of behavior emerge. Moment-to-moment interactions in adult–infant dyads seem to be founded on contingent responsiveness. Parents acknowledge the spontaneous acts of infants and treat them as inherently meaningful. In this work, like we did in previous chapters, we will foreground adult behavior. In this case, we will deal with adult behaviors contingent upon infant cues.

Studies on infant behavior feedback within social interaction are commonly based on the concept of *maternal or parental responsiveness* (both terms are used interchangeably). When infants interact with social partners, there is an almost immediate feedback response. The temporal coordination of behavior is regarded as the basis of social interaction. It can be used to establish the behavior of someone as a response to our own, and thus as self-relevant (Henning & Zmyj, 2016). We will therefore focus on the organized temporal coordination between infant cues and maternal response. Adult feedback toward infant behavior during social interaction is not elicited by every infant behavior. Thus, it is imperfectly contingent upon the type of behavior, and usually involves some uncertainty with respect to what type of infant behavior will be responded, and how likely the assessment of that behavior will be (Henning & Zmyj, 2016). However, maternal actions are thought to drive certain infant conduct. From day one, mothers respond to the signals of their infants; for example, by providing vocal feedback for most of the vocalizations they produce (Dominguez et al., 2016; Feldman & Eidelman, 2007). When responding, they let them know that they have noticed what they just did: "you talk to me," "you listen to me" (Dominguez et al., 2016).

Similarly, infants respond with nonverbal vocalizations that are recognized by parents. Such nonverbal affective transmission rhythmically adjusts calls and responses following an intersubjective dialogue (Volgsten, 2012). In a study with 4-month-old infants, Van Egeren et al. (2001) found that vocalizations tended to be the primary response and the most salient cue for both partners, emphasizing the role of vocal behaviors over other actions. Studies on maternal responsiveness typically highlight vocalization, smiling, and gaze dynamics (Bigelow & Power, 2016;

Nomikou et al., 2016), which are prevalent in face-to-face interaction. This research standpoint is likely guided by the search for cues to language acquisition, characteristic of the *linguistic turn* (Sheets-Johnstone, 2015). We prefer to challenge this perspective by studying the distribution of vocal and kinesthetic behaviors in maternal responsiveness during body-to-body interaction. As noted in Chap. 2, when the emotional expressions, vocalizations, and movements of the infant are followed by the movements, expressive comments, and gestures of the adult, the infant elicits increasing social behaviors, which in turn encourage a response from the adult, resulting in *social circular reactions* (Español, 2007, Gouin Décarie, 1962, Rivière, 1986/2003, and see Chaps. 2 and 9). Social circular reactions are joyful and vital sonorous-kinesthetic encounters. Given the multisensory character of adult behavior during body-to-body encounters, we believe that expanding our focus beyond vocal behavior, to include movement in the study of parental responsiveness, will bring us closer to a better understanding of the phenomenon.

Research on parental responsiveness generally focuses on *what* is shared (e.g., matching affective or other behavioral states) or/and on the temporal structure, that is, the sequence and timing of parent and infant conducts (Henning & Zmyj, 2016). But, in addition, adults do not always respond in the same way: parental responsiveness is rich in variation and includes multiple reciprocity patterns (adjustments of body posture, vocalization, speech, etc.). Within the varied reciprocity patterns shaping the flow of interaction, we find *matching activities*, a type of interactive behavior involving agreement or similarity between the behaviors of both partners. Matching activities set up a special degree of mutuality (Beebe et al., 2003; Užgiris et al., 1989) and social commitment (Agnetta & Rochat, 2004; Bordoni et al., 2016). In this chapter, we focus on two particular matching activities within parental responsiveness: *imitation* and affect *attunement*.

Since his seminal studies, Daniel Stern, 1985 identified a particular kind of matching activity that falls in the same continuum as imitation: affect attunement. In early developmental stages, imitation becomes an interactive resource where child and adult engage in the same behavior. Unlike imitation, in affect attunement, adults respond to infant behavior in a different way but maintain and reflect certain characteristics of the original behavior, such as rhythm or intensity. In contrast to imitation, which is part of the behavioral repertoire of both partners, affect attunement has been described only as part of the repertoire of adult behaviors. This chapter aims to explore two short units of parental responsiveness – imitation and affect attunement – and their temporal coordination.

4.1.3 Parental Imitation and Affect Attunement

Imitation is an interactive resource that allows similarity and mutuality in interpersonal contact from very early stages of development, evidencing the precocious sensitivity and social orientation of infants (Bordoni, 2018a; Devouche & Gratier, 2019; Maratos, 1998; Reddy, 2008a; Vincini et al., 2017). Beyond the current

controversy over the interpretation of neonatal imitation experiments (Keven & Akins, 2017; Meltzoff et al., 2017; Oostenbroek et al., 2016; Vincini et al., 2017), imitation is indeed observed in the dyad later in development. During the first year of life, infants and adults use it to engage in intersubjectivity and extend cycles of continuous interaction (Kugiumutzakis et al., 2005; Nagy & Molnar, 2004; Pawlby, 1977). Early imitative play between adults and infants develops with positive emotions, interest, and enjoyment (Kugiumutzakis et al., 2005; Pratikaki & Kokkinaki, 2013), composing an enjoyable experience that provides an emotional reward to both participants. The imitative reaction of the caregiver becomes a source for infant self-knowledge, giving infants the opportunity to see and objectify their impact on others (Gergely & Watson, 1999; Rochat, 2001). When adults imitate infants, they perceive exteroceptively what they feel proprioceptively, undergoing a particular combination of exteroception-propioception, both fundamental experiences for self-development and social bonding (Agnetta & Rochat, 2004; Español, 2012; Reddy, 2008a).

During early social play, adults can influence and shape infant behavior according to what they consider culturally valuable. Imitation is not arbitrary but selective; the selected event is relevant to both subjects and guides the content of the interaction (Gratier & Devouche, 2011; Užgiris, 1999). Another pattern of reciprocity used by adults to initiate affective communicative coordination with their infant is affect attunement, a phenomenon recognized and defined by Stern in his pioneering work (Stern, 1985; Stern et al., 1985) when describing moment-to-moment interactions within the mother–infant dyad (see Ammaniti, 2018 to review this topic). As we mentioned above, opposed to imitation, which forms part of the behavioral repertoire of adults and infants, studies suggest that at least during the first year of life, affect attunement is practiced only by the adult. It was first observed in spontaneous mother–infant play around the ninth month (Stern et al., 1985), but its early occurrence was recently reported from the second month onward (Español et al., 2018; Jonsson & Clinton, 2006; Jonsson et al., 2001). At first glance, affect attunement might be mistaken for imitation since they fall into the same spectrum but they are different phenomena (Stern et al., 1985). In affect attunement events, adults respond to infants in a different way, while retaining and reflecting certain characteristics of the original behavior. It is a specific way of responding to infant behavior where certain features are selectively imitated and expressed through a different and novel conduct. Briefly, affect attunement is defined as a specific type of adult behavior involving an open-ended conduct different from that of the infant that preserves at least one amodal feature including the intensity (absolute or intensity profile), the temporal pattern (duration, rhythm, pulse), or the spatial pattern (Stern, 1985). These amodal features are peculiar (see Chap. 5) because they can be perceived by several perceptual modalities since they are not specific to a particular one. For example, a 9-month-old baby is very excited when she sees a toy and wants to pick it up. In doing so, she exuberantly utters “Aaaaaaaaaaah” and looks at her mother. She looks back at her, grabs her own shoulders, and twirls her upper body dramatically, like a disco dancer. The dance and expression are equally long, exciting, joyful, and intense. In this case, the infant vocalizes and the mother attunes by

performing a movement for her child that retains identical characteristics of duration and intensity (Stern, 1985; Stern et al., 1985). While in imitation the attention focuses on the form of the external behavior or action, in affect attunement it is set on the subjective state, in the *character of the shared feeling*. It thus turns into a privileged way of communing with internal states and indicating others that an internal state is being shared (Stern, 1985). External matching behaviors may differ in form and manner, but they are interchangeable manifestations of a single, recognizable internal state. Mothers pick up certain features in the expression of their infants and turn them into something else, generating a new open-ended behavior. Affect attunement helps establish analogous interaffective experiences other than physical behavioral interaction (Volgsten & Pripp, 2016). It brings increasing awareness that my partner is feeling something comparable to what I feel (Volgsten, 2012). Affect attunement resembles imitation, social contagion, and empathy while enabling emotional resonance. But it sets itself apart because it recasts emotional experiences in a different expression form that reconfigures subjective states, giving them a referent status and treating open-ended behaviors as one possible expression of that referent (Stern, 1985). Mother and infant amodally recognize form, time, and intensity, evoking reciprocal intersubjective reactions, making amodal perception a crucial factor for maternal affect attunement (Ammaniti, 2018). In theoretical terms, affect attunement has been recognized to play an important role in the establishment of intersubjectivity experiences – both in early caregiver–infant interactions and in psychotherapeutic interactions (Ammaniti, 2018). Also in theoretical terms, it has been related to music development (Volgsten, 2012; Volgsten & Pripp, 2016), body image, and self-esteem (Knafo, 2016). It is also a valued resource for music therapy (Trondalen & Skårderud, 2007), and empirical research on embodied attunement in clinical practice (Ammaniti & Ferrari, 2013; Kykyri et al., 2017). Despite the theoretical depth in which it has been approached and its perceived importance for infant social and communicative development (Beebe et al., 2003; Bornstein & Tamis-LeMonda, 2004; Hobson, 1993; Reddy, 2004; Tomasello, 2019), there is relatively scarce baseline research on affect attunement (Bordoni, 2018b).

The Frequency of Maternal Imitation and Affect Attunement, and Its Relation with Infant Reaction

Affect attunement and maternal imitation do not seem sporadic but relatively frequent behaviors in social exchanges with infants. Originally, Stern et al. (1985) reported that, between the ages of 8 and 12 months, the affect attunement rate is almost 1 per minute, while the imitation rate is 0.37 per minute. According to Jonsson et al. (2001), between the age of 2 and 12 months, the rate of matching activities in adult–infant interactions is 1 coincidental event per minute (imitation or affect attunement). It should be stressed that both studies took place in a laboratory context. In a recent single-case study, Español et al. (2018) explored imitation and affect attunement in the home of infants during early social play moments, finding

that, between the ages of 3 and 6 months, the matching activity rate is around 2 events per minute per infant.

Other studies reveal some discrepancy in the frequency ratio of both behaviors. Jonsson et al. (2001) found that imitation patterns are equally (or more than) frequent than affect attunement and that the frequency ratio between imitation and affect attunement changes as the infant ages. Between the second and sixth months, maternal imitation is more frequent than affect attunement, while in the second half of the first year of life, affect attunement events predominate over imitation events. In contrast, Español et al. (2018) found that, between the age of 3 and 6 months, within maternal behavior, attunement patterns almost duplicate those of imitation and tend to decrease toward the end of the period studied, suggesting this might be reflecting a lapse trend, marked by an incipient transition to secondary intersubjectivity, and the maternal adjustment to the changing interests of the infant, who begins to explore the world of objects. Any discrepancies found between the data might be related to variations in research design and the interactive context analyzed, as well as to the interactive style of the studied dyads (from different cultural backgrounds). Español et al. (2018) reported a single-case study analyzing affect attunement and imitation specifically during early social play episodes, whereas in the other mentioned studies, these matching activities were recorded in more general mother–infant social interaction situations (not exclusively in early social play).

Infants are sensitive to adult imitation; they enjoy and pay special attention when their mother imitates what they just did (Kokkinaki, 2003; Pawlby, 1977). Being imitated apparently establishes a powerful and immediate state of interest, connection, and intimacy (Reddy, 2008a). More specifically, infants reportedly devote greater visual attention to the imitating adult, show a higher frequency of smiles (Bordoni et al., 2016; Field et al., 1985; Jones, 2006), and during the second half of the first year of life, are even able to monitor adult imitation by testing it through behavioral modulation (e.g., by abrupt stops or changes in their behavioral direction) (Agnetta & Rochat, 2004; Meltzoff, 1990; Sauciuc et al., 2020). In addition, the effects of *being imitated* have been recently found to manifest neurophysiologically as well (Saby et al., 2012).

We only found two studies analyzing infant responsiveness to maternal affect attunement. Markova and Legerstee (2006) reported that compared to maternal imitation, maternal attunement shows a higher ratio of mother-directed gazes, more smiles, and more positive vocalizations in infants from 5 to 13 weeks. Bordoni et al. (2016) found that, during the 6- to 12-month period, affect attunement generates less infant visual engagement than imitation. The difference in infant responsiveness might be due to the fact that parental affect attunement was operationalized differently: Bordoni et al. (2016) considered affect attunement like Stern et al. (1985), while Markova and Legerstee (2006) regarded it as a way to globally rate maternal behavior or the segment of social interaction in a *high* or *low attunement* gradient.

Based on Jonsson et al. (2001) and Jonsson and Clinton (2006), Bordoni et al. (2016) suggest that, compared to imitation, affect attunement can function as a reciprocity pattern that enables adults to support infants in their increasing exploration

of the world of objects, incipient in the second half of the first year, and enhanced by motor development. The coincidence between the intensity and temporal patterns of the behaviors involved in affect attunement may generate intersubjective encounters beyond eye contact. From the point of view of adults, affect attunement could be a useful way to accompany infant activities without diverting them from what they are doing. From an infant perspective, affect attunement could provide a companion experience to their activities.

Modality Behavior in Parental Imitation and Affect Attunement

Regarding what kind of conduct is involved in imitation events during the first year of life, studies show that facial expressions, vocalizations, and head movements prevail in the first 6 months (Jones & Yoshida, 2011; Kokkinaki & Vitalaki, 2013; Moran et al., 1987; Užgiris, 1984; Užgiris et al., 1989) while in the second half of the year, manual imitation and actions on objects become noticeable (Masur & Rodemaker, 1999). However, most imitation studies in spontaneous mother–infant social interaction contexts at this developmental stage have focused on the analysis of vocalic imitation (Kokkinaki & Kugiumutzakis, 2000; Kokkinaki & Vasdekis, 2003; Kugiumutzakis, 1993; Malloch & Trevarthen, 2008; Papoušek et al., 2000; Papoušek & PapouŠek, 1989).

In the case of affect attunement, different combinations between adult and infant behaviors are possible. In *transmodal attunements*, the matching takes place when the coincidence is expressed in different behavioral modalities; for example, movement rhythms match vocalization rhythms. In *intramodal attunements*, the same behavioral modality is used through a different open behavior. In *mixed attunements*, a different behavioral modality is added to the first one (Stern et al., 1985). In *imitation*, on the other hand, there is always an intramodal match (movement matches movement, vocalization matches vocalization).

There is limited research concerning which infant behaviors are most attuned and which behavioral modalities are used by adults to attune them. Information referring to the types of affect attunement and the behavioral modalities used are based on the original study by Stern et al. (1985) for the period of 8–12 months. Most of the affect attunements observed in this work were transmodal and mixed, indicating mothers attune affectively and thus the behavioral modality of their infants changes. In Stern et al. (1985), 39% of affect attunements were transmodal, 13% intramodal, and 48% mixed, meaning that 87% of the attunements were made in a modality different than that of the infant. Results also suggest that affect attunements are caused by different infant behaviors and are performed by the mother through different behavioral modalities (vocalization, gesture, head movement, body movement, facial expression, breathing). Despite the relevance of these data, there is no information on the behavioral mode combination of both participants.

A Time Window for Contingent Responsiveness and Behavior Modality

The temporal coordination between infant cues and maternal response is central to parental responsiveness research. Some attempts have been made to find out what is the appropriate latency period for contingency definitions in mother–infant interaction (Bornstein et al., 2015, 2021; Devouche & Gratier, 2019). The time window for contingent responsiveness equals the number of seconds elapsed until *the* response probability reaches a *plateau*. Existing studies about the temporal window analyzed different interactive situations and focused on different adult and infant behaviors. Keller et al. (1999) found that the contingent response between maternal utterances and infant vocalization is about 1 s, regardless of the maternal language and culture. Van Egeren et al. (2001) analyzed maternal verbal, smile, and gaze behaviors, maternal touch and play behaviors, and infant verbal, gaze, and play behaviors. They found a latency of 3 s was most appropriate with 4-month-old infants for contingency definitions in mother–infant interaction. Dominguez et al. (2016) reported that the most appropriate time window for mother–newborn protoconversation social contingency was 1-s. Recently, Boiteau et al. (2021) observed newborn–parent dyads finding them responding to each other within a 1–3-s time window. On the other hand, Gratier et al. (2015) identified that temporal coordination in protoconversations with infants aged 2–5 months involved both *turn-taking* and *overlapping* vocalizations. Overlapping vocalizations were coded when either the mother or infant vocalized on top of the vocalization of their partner. The whole vocalization was coded as overlapping even if it was only partially masked by the vocalization of the partner. The pauses between alternating vocalizations (*switching pause*) could range from a few milliseconds to 3 seconds.

According to the reviewed studies, the time window for contingent responsiveness has been mainly studied in the general context of adult–infant interactions with no specification of the type of exchanges produced and in the specific context of protoconversations. Vocal behavior in protoconversations seems to be the most studied phenomenon. It is clear that despite the relevance of temporal coordination in all kinds of early interactions, the temporal window of specific patterns of adult matching activities, such as imitation and affect attunement, to our knowledge, has not been adequately studied. Although vocalization is the most studied phenomenon in adult–infant interaction, early studies have also considered movement. Beebe et al. (1985) analyzed interpersonal synchronization in vocal and kinesthetic mother–infant interaction. Vocalization and movement were analyzed separately. When analyzing vocalizations, they considered the switching pauses and overlapping vocalizations. In the analysis of movement, they considered on–off and holding movements. The holding movement is not differentiated by whether or not they have switched turns. Intelligibility in adult conversation requires an alternation between speaking and listening, a constraint absent in kinesis, where both speakers can act and control themselves simultaneously, while not experiencing the interruption this situation would entail in a verbal exchange. Hence, a distinction is made between *coactive* and *non-coactive turns*. Turns can shift following either joint silence (a switching pause) or joint inaction (holding); in this case, they alternate

smoothly so that one partner acts and holds, and then the second acts and holds. This is a *non-coactive* or alternating turn, evolving from joint silence and prevailing in adult speech. In the other case, the partners join in simultaneously after an action taken by the leader that keeps going once the leading partner finishes. This is a simultaneous switching, a co-active turn exchange. Unlike conversation, this type of turn prevails in mother–infant kinetic interaction. Co-action is a basic aspect of this phenomenon that does not disrupt ongoing communication.

Our aim is to analyze temporal coordination in two specific patterns of maternal behavior: imitation and affect attunement. Given that imitation can involve vocalization/vocalization or kinetic/kinetic pairings and that affect attunement implies a match between different modalities – e.g., vocal/kinetic or kinetic/vocal – we need to analyze the behavioral matching of different modalities. To avoid the multiplication of terms, we will use *overlapping behavior* and *switching pause in alternating behavior* to refer to the coactive and non-coactive temporal coordination between infant behavior and maternal matching, whether kinesthetic, vocal, or a combination of both.

4.2 An Empirical Study on Maternal Imitation and Affect Attunement

In this section, we present a single case study on two patterns of maternal responsiveness: imitation and affect attunement. The aim is to analyze the development of imitation and affect attunement frequencies, the temporal coordination, and the behavioral modalities of both parental responsiveness patterns in the context of early social play, between the second and tenth month of life. The data presented in this section extend those reported in a previous work (Español et al., 2018) involving a mother–infant dyad. For that study we analyzed maternal behavior between the second and sixth month of her baby. In this work, we extend the developmental period studied in the same dyad, to the tenth month, incorporating the analysis of temporal coordination.

4.2.1 Method

Participants

A mother–infant dyad participated in this study. In the first session, the age of the infant – according to the Piagetian convention [year; month (day)] – was 00; 2 (20) in the first session and 00; 10 (21) in the last one. The infant is the second child of a middle-class Argentine family from the city of Buenos Aires. He was born full term and followed standard developmental milestones, as reported to the parents during periodic visits to the pediatrician. Both parents were highly educated.

Procedure

The dyad was always visited at home. According to Bruner (1983), the dyad is sensitive to the environment, which makes the natural messiness of home life preferable to the one artificially created in a laboratory. In this environment, communication will likely be motivated differently than in a lab setting, with preset dispositions for registering or testing reactions in staged performances (Beebe et al., 1985). The mother was asked to engage in normal interaction with her baby as she does in her daily life.

Fifteen sessions of approximately 45 min each (at 15-day intervals) were filmed in their home, registering about 675 min of spontaneous interaction, which included all kinds of situations (feeding, bathing, diaper changing, cooing, pain soothing, and social play).

Infant–adult interaction involves a wide range of engagement and, as mentioned in the introduction, reciprocity within the dyad is structured through short units of interactive behaviors that may or may not be organized in extended temporal sequences. In the present study, two brief patterns of maternal matching responsiveness (affect attunement and imitation) and their temporal coordination were studied in situations of especially involved interaction: early social play. Early social play often unfolds in the midst of other forms of interaction. While their distinction is not always evident, a primacy of means over ends and a nonfunctional orientation are two constitutive features that some consider sufficient to characterize this play (Pellegrini, 2009; Fagen, 2011). Early social play is a particularly joyful form of interaction whose sole purpose is social exchange and enjoyment (Stern, 1974, Whaley, 1990, Chaps. 2 and 9 in this volume). It is pure social interaction that once initiated stops all other external tasks while interpersonally focused events come to the fore (Stern, 1985). During early social play, when the dyad is observed, infants demonstrate an active emotional initiative with a great variety of expressions, and parental responsiveness becomes richer.

To analyze maternal matching responses in early social play, we selected the first 10 min (not necessarily continuous) of early social play from each session. The detection of early social play situations met three criteria: adult and infant were socially engaged (any of the following indicators of social engagement were observed: mutual gaze and/or smiling and/or seeking body contact); the engaged behavior should lack a functional orientation (meeting the basic needs of the infant); and a primacy of means over ends should be achieved through a pleasurable repetition of the nonfunctional behavior. Based on this selection of interactive situations, a 150 min *developmental video* was made by selecting the first 10 min (not necessarily consecutive) of each early social play session. Developmental videos are a crucial step in change process research: they collect any relevant information in a relatively short documentary film, facilitating greatly their iterative review (Fogel & DeKoeyer-Laros, 2007).

Observation Code

Three dimensions of adult responsiveness were studied: (i) the pattern of maternal matching response, (ii) the behavioral modality of infant signals and maternal response, and (iii) the type of temporally established coordination. The categories through which dimensions (i) and (ii) were analyzed are the same used by Español et al (2018), while the categories corresponding to dimension (iii) were taken from Gratier et al. (2015)

4.3 Patterns of Maternal Matching Response

(a) *Imitation*: The mother responds to the immediate preceding behavior of the infant by globally matching the action of the infant; she responds in the same behavioral modality with the same action content. For example, if the infant opens the mouth, she opens her mouth (imitation of facial expression); if the infant vocalizes, the mother vocalizes using the same utterance (vocal imitation).
(b) *Affect attunement*: The mother responds to the immediate preceding behavior of the infant through a behavior, different from that of the infant, by matching the duration, rhythm, pulse, quantity, intensity, and/or spatial pattern of behaviors. In *transmodal affect attunement*, her behavioral modality is completely different (e.g., when the infant nibbles or sucks on her cheek, she vocalizes to the rhythm and intensity of the mouth and tongue movements felt on her face). In *intramodal affect attunement*, the mother attunes to the same behavioral modality (e.g., when the infant vocalizes "Aaa aaa aaa," the mother responds with the same intensity and rhythm "Mmm mmm"). In *mixed affect attunements*, the mother may add an additional modality to the same behavioral modality (for example, in addition to attuning by vocalizing to the strenuous behavior of the infant, the mother adds a facial expression of effort by closing her eyes and mouth).

4.4 Behavioral Modality of Infant Signal and Maternal Response

(a) *Vocalization:* Includes any vocally emitted sound, from syllabic or near-verbal vocalizations to grunts and onomatopoeias.
(b) *Facial expression:* Includes motional expressions and exaggerated or unusual facial gestures. For example, opening the mouth, closing the eyes or blinking exaggeratedly, sticking out the tongue, etc. Smiles were excluded from coding, as they were very frequent and accompanied many situations during interactions.

(c) *Movement:* Includes all bodily movements other than facial expressions, those that can be seen (proximal and distal), and those that are only perceived proprioceptively (when the spatial positions of mother and infant block the visual perception of movement).

Infant behavior is often multimodal. For simplicity, we decided to categorize the prevailing or guiding interaction modality. For example, in the case of infant vocalizations, which evidently involve facial expression, the behavior was categorized as vocal, and we restricted the category to facial expression or movement when these occur in the absence of vocalization or are the most conspicuous behavior.

4.5 Temporal Coordination Between Infant Signal and Maternal Response

(a) *Switching pause in alternating behaviors:* It refers to the pause between the infant signal (facial expression, vocalization, or movement) and the maternal response (facial expression, vocalization, or movement). Following the criteria used by Gratier et al. (2015) in their turn-taking study on mother–infant vocal interaction, switching pauses might oscillate between a few milliseconds and 3 seconds.
(b) *Overlapping behavior*: It was coded when the maternal behavior (facial expression, vocalization, or movement) was superimposed on the infant signal (facial expression, vocalization, or movement). Similarly to Gratier et al. (2015), we coded the entire behavior as overlapping even when it was only partially masked by maternal behavior.

4.5.1 Analytic Strategy and Coding Reliability

The coding categories and the developmental video were incorporated into the Anvil 5.1.9 (Kipp, 2008) video annotation software. Based on the code applied in the developmental video, each maternal matching response event (imitation and affect attunement) was identified. Affect attunement was classified by type: transmodal, intramodal, or mixed. For each imitation and affect attunement event, we analyzed the behavioral modality of maternal responses (facial expression, vocalization, movement), the modality of infant signals that elicited them (facial expression, vocalization, movement), and the correspondence between both modalities. The frequency of maternal matching responses was calculated by counting the total number of imitation and affect attunement events per session. The proportion of each of the imitation and affect attunement behaviors was calculated as a percentage. The matching response rate was calculated by dividing the number of events

identified by the total observation time (150 min). Finally, we analyzed the temporal micro-organization of maternal matching responses (overlapping or alternating).

The coding process involved two partners (the first four authors of this article), followed by a discussion on any disagreement points. We followed these steps: (a) each recorded session was thoroughly observed by two coders; (b) maternal matching responses subject to disagreement between coders were discussed by the whole group until a consensus was reached; and (c) in the few occasions where consensus was not reached, the event was dismissed. This procedure is consistent with those used in social interaction studies (Español et al., 2015; Español et al., 2018; Forster & Iacono, 2013; Jonsson et al., 2001). The selection of early social play episodes composing the developmental movie was conducted in the same way.

4.5.2 Results

How Often Does the Mother Attune or Imitate Infant Signals When They Play Together?

Within the 150 min of analyzed early social play, we identified a total of 341 maternal matching response events. Table 4.1 shows that 221 episodes were affect attunements (65%) and 120 episodes were imitations (35%). The frequency of maternal affect attunements ranged from 2 to 31 events per 10 min of early social play, while maternal imitations offered less variability (ranging from 2 to 15 events per session). During the studied period, the frequency of affect attunement events exceeded the frequency of imitation events in most of the sessions. Only in three opportunities – at ages 00; 02 (20); 00; 03 (05); and 00; 06 (20) – the frequency of imitation events exceeded that of affect attunement events.

Out of 221 affect attunement events, 68% were transmodal, 20% mixed, and 12% intramodal. Table 4.1 also shows the distribution by infant age of maternal affect attunement types during the studied period. The mother engaged in transmodal affect attunements at all times, with a frequency ranging from 2 to 24 events every 10 min of early social play. In contrast, intramodal and mixed events only occurred in some sessions.

To date, we found only one study that reports the distribution by type of affect attunement (Stern et al., 1985), indicating that, for infants aged 8 to 12 months, the distribution of affect attunements was 13% intramodal, 39% transmodal, and 48% mixed. When grouping our data in periods 2 to 7 months and 8 to 10 months, the distribution of attunement types differs between periods (see Table 4.2). In the first period (2–7 months), the distribution was 13% intramodal, 7% mixed, and 80% transmodal, while in the period 8–10 months it was 11% intramodal, 48% mixed, and 41% transmodal. The difference we found between both periods was statistically significant for mixed ($X^2 = 48.875$, $p = 0.000$) and transmodal ($X^2 = 33.635$, $p = 0.000$) attunements. Thus, the data collected in our study concerning the distribution of affect attunement types in the period 8–10 months were similar to those

Table 4.1 Frequency of maternal matching response events

Age	Maternal Matching Response				
	Imitation	Affect Attunement	Intramodal Affect Attunement	Mixed Affect Attunement	Transmodal Affect Attunement
00; 02 (20)	13	11	1	2	8
00; 03 (05)	10	2			2
00; 03 (23)	10	14	2	2	10
00; 04 (09)	9	17	1		16
00; 04 (28)	14	29	2	3	24
00; 05 (21)	4	16			16
00; 06 (02)	8	23	4	2	17
00; 06 (20)	8	6	3		3
00; 07 (03)	2	6			6
00; 07 (14)	4	17	5	2	10
00; 07 (28)	8	9	1		8
00; 08 (20)	6	20	1	6	13
00; 09 (11)	2	10		5	5
00; 09 (26)	7	10	2	3	5
00; 10 (21)	15	31	5	20	6
Total	120	221	27	45	149

Frequency of maternal matching response events (imitation and affect attunement) every 10 minutes of early social play by infant age [year; month (days)], during the period from 2 to 10 months. In the shaded columns, we present the frequency detail of each type of affect attunement (intramodal, mixed, and transmodal)

found by Stern et al. (1985) in their original study and different from the ones reported for the preceding developmental period (2–7 months).

In this type of study, a further way to weigh how often do mothers attune or imitate infant cues is by calculating the occurrence rate of maternal matching activity events (Español et al., 2018; Jonsson et al., 2001; Stern et al., 1985). Over the

Table 4.2 Distribution of affect attunement types

	2–7 months	8–10 months
Intramodal affect attunement	13%	11%
Mixed affect attunement	7%	48%
Transmodal affect attunement	80%	41%

Distribution (in %) of affect attunement types for periods 2–7 months and 8–10 months

Table 4.3 Distribution of type of maternal matching response based on the modality of infant signals

		Maternal matching response	
		Imitation	Affect attunement
Infant signal	Vocalization	79%	19%
	Movement	7%	78%
	Facial expression	14%	3%

Distribution (expressed in percentages) of the type of maternal matching response (imitation and affect attunement) based on the modality of infant signals (facial expression, movement, and vocalization) in early social play. Period 2–10 months

observed period, the event rate of matching activities was 2.27 events per minute (341/150); the rate of affect attunement was 1.47 (221/150) per minute, and the rate of maternal imitation was 0.80 (120/150) per minute.

How Does the Mother Respond to the Signals Offered By Her Infant?

As shown in Table 4.3, maternal imitations were mainly caused by baby vocalizations (79%), followed by facial expressions (14%) and, finally, by movements (7%). In contrast, maternal affect attunements were mainly elicited by baby movements (78%), then by vocalizations (19%) and, finally, by facial expressions (3%). The maternal tendency to imitate the vocalizations and to attune to the movements of her infant was significant ($X^2 = 160.32$, $p = 0.000$).

What Goes with What? Matching Between Infant and Maternal Behavioral Modalities

We observed all maternal affect attunement and imitation events with their corresponding matched behavioral modalities, i.e., with which modality (vocalization, facial expression, and movement) the mother responded to the different infant signals (vocalization, facial expression, and movement) (see Table 4.4). In terms of the behavioral modalities of maternal response based on the behavioral modalities of the infant, we found a significant difference ($X^2 = 213.118$, $p = 0.000$). Results indicated that infant movement signals (182 events during the 2- to 10-month period)

were mainly responded with vocalizations (90%), either single (71%) or combined with movement (19%); to a lesser extent, with movements (7%), and finally with facial expressions (3%). Infant vocalizations (136 events) were also mainly responded with vocalizations (90%), significantly less with facial expressions (8%) and movement (2%). Finally, infant facial expressions (23 events) were mainly responded with the same behavioral modality, either singly or combined with vocalizations.

How Is Maternal Matching Responsiveness Manifested? Temporal Coordination and Matching Modalities in Affect Attunement and Maternal Imitation

For each maternal affect attunement and imitation event, the temporal distance between the maternal behavior and the infant behavior triggering it (regardless of the type of behavior involved: vocalization, facial expression, or movement) was observed. Out of 341 maternal imitation and affect attunement events, 208 were overlapping behaviors (61%) and 133 were alternating behaviors (39%). For these 133 events, alternating pause lengths were mostly equal to or less than 1 second. In only 5 events, the length was between 1.24 and 2.6 seconds. Out of 120 imitations, 74% took place in a switching pause (89 events) and the remaining 26% overlapped (31 events), whereas out of 221 affect attunements, 20% occurred in a switching pause (44 events) and 80% overlapped (177 events). The difference was statistically significant ($X^2 = 96, 23$, $p = 0.000$).

The most frequent combination in maternal matching responsiveness was overlapping affect attunement responding to infant movement. Maternal imitation in response to infant vocalizations in turn-taking, with a switching pause was the next most frequent combination. The frequency distribution of maternal matching responsiveness events based on the behavioral modality of infant signals (facial expression, movement, or vocalization), the maternal matching response type (affect attunement and imitation), and the temporal organization type of the exchange (overlapping or alternating) are shown in Table 4.5.

Table 4.4 Distribution of mother response modalities based on the modality of infant signals

		Infant signals		
		Movement	Vocalization	Facial expression
Mother response	Movement	7%	2%	4%
	Vocalization	90%	90%	9%
	Facial expression	3%	8%	87%

Distribution (in percentages) of maternal matching response behavior modalities (vocalization, movement, or facial expression), based on the behavior modality of infant signals (vocalization, movement, or facial expression)

4.5.3 Discussion

The present study was designed to provide a comprehensive description of naturalistic patterns of maternal responsiveness during the period between the second and tenth months of infant life in the context of early social play. It focused on maternal affect attunement and imitation, characterized by the creation, on the part of the mother, of a match between her behavior and that of her infant.

The social contingency time window for matching activities. The studied matching activities materialize in a short time window – mostly equal to or less than 1 second. The time window turned out to be shorter than the one suggested by Van Egeren et al. (2001) and Boiteau et al. (2021), and considerably similar to the one proposed in neonatal protoconversation studies (Dominguez et al., 2016). This finding reflects how readily the mother is able to respond to her infant using matching activities in the context of early social play. Imitation or affect attunement is an intuitive and very accessible reciprocity pattern within the organization of early social play. Since this is a single case study, this interesting result should be corroborated in future research.

The overlapping and alternating behaviors. More than half (61%) of the maternal response pattern events were synchronous with infant signals, which means that they were overlapping behaviors. The remaining 39% were alternating behaviors within switching pauses equal to or less than 1 second. These results highlight that in early social play, the matching activities of maternal responsiveness are temporally coordinated in the setting of very short time periods. Since this time window remains constant throughout the period, the temporality in which the whole interaction occurs also remains constant.

The event rate of matching activities. In the 2–10 month period, the rate of matching activity events was approximately 2 per minute of social play, exceeding the rate found by Stern et al. (1985) for the 8–12 month period and by Jonsson et al. (2001) for the 2–12 month period, which was approximately 1 coincidence event per minute. It is consistent with the results of Español et al. (2018) for the 3–6 months period. The differences among these data may be due to the diversity in research design and the interactive context analyzed (spontaneous social interaction vs. early

Table 4.5 Distribution of the frequency of overlapping or alternating maternal matching response

		Facial expression	Movement	Vocalization	Total
Imitation	Overlapping	12	7	12	31
	Switching pause	5	1	83	89
Affect attunement	Overlapping	4	164	9	177
	Switching pause	2	9	33	44
Total		23	181	137	341

Distribution of the frequency of maternal matching response events according to the behavioral modality of infant signal (facial expression, movement, or vocalization), to the type of maternal matching response (affect attunement and imitation), and the type of temporal organization of the exchange (overlapping or alternating)

social play). Beyond these differences, all these studies clearly show that matching activities are not sporadic but frequent events in spontaneous mother–infant social interaction, and even more so in the context of early social play. As we pointed out in the introduction, one of the characteristic features of early social play is a high level of social involvement between the infant and the adult (Stern, 1974; Whaley, 1990). The high rate of matching activities events may be considered as a corroboration of the intense dyad involvement in early social play contexts.

The frequency of maternal affect attunement and imitation. Affect attunement responses were more frequent than imitations (65% versus 35%), consistent with the results found by Español et al. (2018) for a shorter observation period, but differing from Jonsson et al. (2001). These differences may be due to the same reasons pointed out previously. However, according to our results, in early social play contexts, affect attunement outperforms maternal imitation – the most studied matching activity pattern in developmental psychology. This is a clear indicator of the relevance of affect attunement in early interactions. Further studies are needed to confirm these findings.

The developmental course of maternal imitation and affect attunement. The data confirm the early existence (prior to the third month) of maternal imitation and affect attunement previously reported by Jonsson et al. (2001) and Español et al. (2018). They also indicate a frequency peak in both patterns in the fourth, fifth, and tenth month. Initially (at 2 and 3 months), imitation events outnumbered affect attunements, a trend that reversed in the following months. A slight decrease in both patterns was detected toward the end of the sixth and beginning of the seventh month, reversing to increase during the following period. This rules out the hypothesis of Español et al. (2018), proposing that a decline around the seventh month might be indicating a fading tendency in both patterns, marking the transition to secondary intersubjectivity. Conversely, these maternal responses peak around the tenth month. However, this peak might still be linked to the transition to secondary intersubjectivity: it may be that joint attention interactions (prototypical of secondary intersubjectivity) expand maternal opportunities to reflect on infant responsiveness and shared intentional states. Further studies are needed to explore this hypothesis.

The recurrence of maternal affect attunements over the extended developmental span studied indicates the relevance of this reciprocity pattern. To a lesser extent, imitation is also frequent throughout the studied period. Evidently, imitation is an accessible interactive resource that might be functioning as a reliable foundation for reciprocity. Both patterns are present throughout, becoming more frequent toward the end of the studied period. In early social play contexts, they emerge as a growing constant in mother–infant interactions during the first year of life. Stern (2010) hypothesized that certain basic reciprocity modes remain present and stable from the beginning of life (in early social play) to adulthood, especially in the temporal/performative arts like music, dance, and theater (see Chaps. 2 and 9). Despite changes inherent to this stage in development (the passage from primary to secondary intersubjectivity, the beginnings of intentional communication, among others), our results on the temporal coordination and frequency favor his hypothesis.

Additional studies are needed to investigate the presence of these reciprocity patterns later in life to prove their continuity or transformation in the development of adult–infant interaction.

Transmodal, intramodal, and mixed affect attunements. All three types of affect attunement described by Stern (1985) were used by the mother. Their distribution was similar to that found by Stern et al. (1985) in his pioneering work. The most frequent responses were transmodal attunements (68%), followed by mixed attunements and very scarce intramodal attunements. The results – confirming the findings of Stern et al. (1985) – indicate that the mother predominantly attunes affectively, changing the behavioral modality of her child throughout the studied period. Our study also provides new evidence on the development and distribution of affect attunement in the first year of life during early social play: transmodal affect attunements were present throughout, while mixed and intramodal attunements emerged only in some sessions, particularly in the last one (around the tenth month).

By segmenting our data into two periods, we identified that the distribution of affect attunement types corresponding to the period 8 to 10 months is very similar to the one reported by Stern et al. (1985) and that, in turn, this distribution differs from the one corresponding to the period 2 to 7 months. To the best of our knowledge, this is the first reported evidence on the distribution of attunement types during the first months of life. These differences refer us to the change process in the development of infant abilities. On the one hand, they are coincident with the so-called periods of primary and secondary intersubjectivity, and on the other hand, with some important movement milestones such as sitting up and crawling. It can be hypothesized that the mother intuitively accompanies and scaffolds these developmental transitions, adapting her own behavior to what she observes in her child.

The modality of maternal response to infant signals. The most likely maternal response to infant signals was vocalizations (90%), with occasional movements or facial expressions. Furthermore, infant vocalizations were mostly responded with vocalizations (90%), the rest being movements and facial expressions; infant facial expressions were mostly responded in the same behavioral modality, often combined with vocalizations. Cross-cultural research on parenting shows how the belief and value systems that define different cultures influence the specific ways in which parents provide companionship and infant care (Papoušek & Papoušek, 2002; Bornstein, 2012). Culture-specific differences in contingent modality response patterns emerge during the second and third months of life (Kärtner et al., 2010). Regardless, vocalizations seem to be the main response tendency of caregivers as well as the most frequent and salient infant signal across sociocultural contexts (Kärtner et al., 2008). The maternal tendency to respond with vocalizations to infant signals found in our study is consistent with these results. Additional cross-cultural studies are needed to explore this tendency.

Temporal coordination between infant behavior and maternal matching. Although maternal imitation and affect attunement are matching activities, they exhibited different temporal coordination patterns. While imitations were predominantly coordinated in turn alternation with a switching pause, affect attunements

were coordinated overlapped, that is, within simultaneous and overlapping infant–mother behaviors. The differences between both patterns were statistically significant; however, the limits of statistical analysis in case studies must be considered.

Until now, it was only known that affect attunements were elicited by infant behavior and performed by the mother along different modalities. No information has been available so far on the combination of mother–infant behavioral modalities. We can now specify that infant movements elicited maternal affect attunements in the first place (78%), followed by vocalizations and facial expressions. On the other hand, during early social play, maternal imitations were elicited mainly by infant vocalizations (79%), followed by facial expressions, and scarcely by movements. Statistical analyses showed a significant tendency in maternal responsiveness to imitate infant vocalizations and to attune infant movements. However, once again, we must emphasize the limits of statistical analysis in case studies. In any case, these data could explain the few existing research works on affect attunement by developmental psychology influenced by the linguistic turn (Sheets-Johnstone, 2015), as well as the current renewed interest in affect attunement in embodied disciplines, where movement takes a relevant place (Durt et al., 2017; Fusaroli et al., 2009; Trevarthen, 2009), because addressing affect attunement requires addressing movement, at least in infant behavior. Likewise, subject to generalization limits imposed by case studies, our work shows that mothers are not only contingently responsive to infant vocalization, smile, and gaze behaviors (as indicated by Bigelow, 1998, Chow et al., 2010, Jaffe et al., 2001, Keller et al., 1999, Malatesta & Haviland, 1982, Stern, 1974, Van Egeren et al., 2001), but also to whole-body infant movement.

*Affect attunement/overlapped/*in response to infant *movement,* and *maternal imitation/in a switching pause/*in response to infant *vocalization.* This study also offers a close look into the construction of circular social reactions made by the dyad in moment-to-moment reciprocity, through the combination frequency of infant behavioral modalities (facial expression, movement, or vocalization), the type of maternal matching response, and the type of temporal organization (overlapped or with a switching pause in alternating behaviors). The most frequent combination was overlapped affect attunement in response to infant movement. This combination reveals a maternal attentive reading of infant movement (even when responding in a different modality), with synchrony being the privileged temporal coordination mode in the kinesthetic domain. Reading these results in light of the fact that the mother was responding primarily with her voice to the movement of her infant, we can surmise that when she was responding with transmodal affect attunements as she vocalized, she was *musicalizing* the movement of her infant.

The next most frequent combination was maternal imitation – performed with a switching pause – in response to infant vocalizations. This combination stresses the switching pause as a privileged mode of temporal coordination in vocal exchanges. The paucity of movement imitation may indicate this reciprocity pattern is not an accessible resource for adults to interact with infants under one year of age. In part, this may be due to the motor differences between them and the supportive role

played by adults, who use their bodies to comfort infants and leave the voice readily available to respond.

4.5.4 Final Remarks

Our study provides new information about two recognized matching activities that establish a particular degree of mutuality or intersubjectivity: maternal imitation and affect attunement.

First, we show that between the ages of 2 and 10 months, over half of all maternal matching responses emerged in synchrony with infant cueing, and alternation occurred with switching pauses shorter than 1 second. This time window remained constant during the studied period. These findings highlight that in early social play, maternal responsiveness matching activities are temporally coordinated according to very brief periods of time.

In the second place, our results indicate that affect attunements were more frequent than imitations. It should be noticed that the frequency of affect attunements exceeds the frequency of the most studied matching activity pattern in developmental psychology. It clearly indicates a need for further empirical studies on affect attunement. Third, they offer a close look at the construction of the circular social reaction of the dyad in moment-to-moment reciprocity, through the frequency of combinations of infant behavioral modalities (facial expression, movement, or vocalization), the type of maternal matching response, and the type of temporal organization (overlapped or with a switching pause in alternating behaviors). Two modes of coordination, combinations, and relationships were described: overlapped affect attunements in response to infant movement; and maternal imitation with a switching pause in response to infant vocalizations. These two forms continued throughout the studied period, revealing a distinctive way of occurrence of every maternal response pattern studied.

Fourth, it highlights that affect attunement patterns grant the creation of intersubjective encounters beyond eye contact that are mainly based on maternal receptivity and reciprocity of infant movements. Affect attunement leads the focus of attention to subjective states, to the nature of shared feelings (Stern, 1985), and favors a peculiar intersubjective encounter: the experience of companionship (Bordoni et al., 2016). This experience of togetherness enables the adult to accompany – not teach, indicate, or impose – the movements and actions of the infant without distracting him/her from what he/she is experiencing. At the same time, it provides the baby with a sensitive, noncoercive, and nonmodeling experience of accompaniment, which may prove to be cardinal for a calm and secure self-development. Unlike cross-sectional studies, in single-case studies the dyad can be meticulously followed in its full dimension. However, they must not be used to make generalizations or establish substantive explanations. Future studies could test these conclusions, which are the result of the detailed nature of single-case research.

References

Agnetta, B., & Rochat, P. (2004). Imitative games by 9-, 14-, and 18-month-old infants. *Infancy, 6*(1), 1–36. https://doi.org/10.1207/s15327078in0601_1

Ammaniti, M. (2018). Implicit knowledge from infancy to the psychotherapeutic relationship: The contribution of Daniel Stern. *Psychoanalytic Inquiry, 38*(2), 138–147. https://doi.org/10.1080/07351690.2018.1405670

Ammaniti, M., & Ferrari, P. (2013). Vitality affects in Daniel Stern's thinking—A psychological and neurobiological perspective. *Infant Mental Health Journal, 34*(5), 367–375. https://doi.org/10.1002/imhj.21405

Beebe, B., Feldstein, S., Jaffe, J., Mays, K., & Alson, D. (1985). Interpersonal timing: The application of an adult dialogue model to mother-infant vocal and kinesic interactions. In T. M. Field & N. Fox (Eds.), *Social perception in infants* (pp. 217–247). Ablex Publishing.

Beebe, B., Rustin, J., Sorter, D., & Knoblauch, S. (2003). An expanded view of intersubjectivity in infancy and its application to psychoanalysis. *Psychoanalytic Dialogues, 13*(6), 805–841. https://doi.org/10.1080/10481881309348769

Beebe, B., Messinger, D., Bahrick, L. E., Margolis, A., Buck, K. A., & Chen, H. (2016). A systems view of mother–infant face-to-face communication. *Developmental Psychology, 52*(4), 556–571. https://doi.org/10.1037/a0040085

Bigelow, A. E. (1998). Infants' sensitivity to familiar imperfect contigencies in social interaction. *Infant Behavior and Development, 21*(1), 149–162. https://doi.org/10.1016/S0163-6383(98)90060-1

Bigelow, A. E., & Power, M. (2016). Effect of maternal responsiveness on young infants' social bidding-like behavior during the still face task. *Infant and Child Development, 25*(3), 256–276. https://doi.org/10.1002/icd.1974

Bigelow, A. E., & Rochat, P. (2006). Two-month-old infants' sensitivity to social contingency in mother–infant and stranger–infant interaction. *Infancy, 9*(3), 313–325. https://doi.org/10.1207/s15327078in0903_3

Boiteau, C., Kokkinaki, T., Sankey, C., Buil, A., Gratier, M., & Devouche, E. (2021). Father–newborn vocal interaction: A contribution to the theory of innate intersubjectivity. *Infant and Child Development, 30*(5), e2259. https://doi.org/10.1002/icd.2259

Bordoni, M. (2018a). La imitación reconsiderada: su función social en la infancia temprana [Imitation reconsidered: Its social function in early infancy]. *Interdisciplinaria, 35*(1), 119–136. https://doi.org/10.16888/interd.2018.35.1.6

Bordoni, M. (2018b). El entonamiento afectivo en las interacciones tempranas adulto-bebé: Una revisión. [Affect attunement in early adult-infant interactions: A review]. *Revista Colombiana de Psicología, 27*(1), 13–25. https://doi.org/10.15446/rcp.v27n1.61019

Bordoni, M., Español, S., & de Grande, P. (2016). La incidencia del entonamiento afectivo y la imitación en el involucramiento visual-social temprano [The incidence of affect attunment and imitation in early visual-social engagement]. *Avances en Psicología Latinoamericana, 34*, 487–503. https://doi.org/10.12804/apl34.3.2016.04

Bornstein, M. H. (2012). Cultural approaches to parenting. *Parenting, 12*(2–3), 212–221.

Bornstein, M. H., & Tamis-LeMonda, C. (2004). Mother-infant interaction. In G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 269–295). Blackwell publishing Ltd.

Bornstein, M. H., Putnick, D. L., Cote, L. R., Haynes, O. M., & Suwalsky, J. T. (2015). Mother-infant contingent vocalizations in 11 countries. *Psychological Science, 26*(8), 1272–1284. https://doi.org/10.1177/0956797615586796

Bornstein, M. H., Manian, N., & Henry, L. M. (2021). Clinically depressed and typically developing mother–infant dyads: Domain base rates and correspondences, relationship contingencies and attunement. *Infancy, 26*(6), 877–900. https://doi.org/10.1111/infa.12427

Bruner, J. (1983). *Child's talk. Learning to use language [El habla del niño (R. Premat, trad.; I. Vila, sup.). Barcelona: Paidós, 1986]*. W.W. Norton.

Chow, S.-M., Haltigan, J. D., & Messinger, D. S. (2010). Dynamic infant–parent affect coupling during the face-to-face/still-face. *Emotion, 10*(1), 101–114. https://doi.org/10.1037/a0017824

Devouche, E., & Gratier, M. (2019). The beginning of parent-infant communication. In G. Apter, E. Devouche, & M. Gratier (Eds.), *Early interaction and developmental psychopathology* (pp. 21–33). Springer. https://doi.org/10.1007/978-3-030-04769-6_2

Dominguez, S., Devouche, E., Apter, G., & Gratier, M. (2016). The roots of turn-taking in the neonatal period. *Infant and Child Development, 25*(3), 240–255. https://doi.org/10.1002/icd.1976

Durt, C., Tewes, C., & Fuchs, T. (Eds.). (2017). *Embodiment, enaction, and culture: Investigating the constitution of the shared world*. MIT Press.

Español, S. (2007). Time and movement in symbol formation. In J. Valsiner & A. Rosa (Eds.), *The Cambridge handbook of socio-cultural psychology* (pp. 238–255). Cambridge University Press.

Español, S. (2012). Semiosis y desarrollo humano [Semiosis and human development]. In J. A. Castorina & M. Carretero (Eds.), *Desarrollo cognitivo y educación: Vol.1. Los inicios del conocimiento [Cognitive development and education: Vol.1. The beginnings of knowledge]* (pp. 219–240). Buenos Aires.

Español, S. (2014). La forma repetición-variación: una estrategia para la reciprocidad [The repetition-variation form: a strategy for reciprocity]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana [Music and developmental psychology. An interdisciplinary exploration of human musicality]* (pp. 157–192). Paidós.

Español, S., Bordoni, M., Martínez, M., Camarasa, R., & Carretero, S. (2015). Forms of vitality play and symbolic play during the third year of life. *Infant Behavior & Development, 40*, 242–251. https://doi.org/10.1016/j.infbeh.2015.05.008

Español, S., Bordoni, M., Carretero Pérez, S., Martínez, M., & Camarasa, R. (2018). La imitación y el entonamiento afectivo en el juego social temprano [Imitation and affect attunement in early social play]. *Interdisciplinaria, 35*(2), 291–305. https://doi.org/10.16888/interd.2018.35.2.3

Fagen, R. M. (2011). Play and development. In A. D. Pellegrini (Ed.), *The Oxford handbook of the develompment of play* (pp. 83–100). Oxford University Press.

Feldman, R., & Eidelman, A. I. (2007). Maternal postpartum behavior and the emergence of infant-mother and infant-father synchrony in preterm and full-term infants: The role of neonatal vagal tone. *Developmental Psychobiology, 49*, 290–302. https://doi.org/10.1002/dev.20220

Field, T., Guy, L., & Umbel, V. (1985). Infants' responses to mothers' imitative behaviors. *Infant Mental Health Journal, 6*(1), 40–44. https://doi.org/10.1002/1097-0355(198521)6:1%3C40::AID-IMHJ2280060107%3E3.0.CO;2-L

Fogel, A., & DeKoeyer-Laros, I. (2007). The developmental transition to the secondary intersubjectivity in the second half year: A microgenetic case study. *Journal of Developmental Processes, 2*, 63–90.

Forster, S., & Iacono, T. (2013). The nature of affect attunement used by disability support workers interacting with adults with profound intellectual and multiple disabilities. *Journal of Intellectual Disability Research, 58*(12), 1105–1120. https://doi.org/10.1111/jir.12103

Fusaroli, R., Demuru, P., & Borghi, A. M. (2009). The intersubjectivity of embodiment. *Cognitive Semiotics, 4*(1), 1–5. https://doi.org/10.1515/cogsem.2012.4.1.1

Gergely, G., & Watson, J. (1999). Early socio-emotional development: Contingency perception and the social-biofeedback model. In P. Rochat (Ed.), *Early social cognition* (pp. 101–136). Erlbaum Publishers.

Gomila, A. (2002). La perspectiva de segunda persona de la atribución mental [The second-person perspective on mental attribution]. *Azafea: Revista de Filosofía, 4*, 123–138. https://doi.org/10.14201/3719

Gouin Décarie, T. (1962). *Intelligence et affectivité chez le jeune enfant* [Intelligence and affectivity in early childhood]. Delachaux & Niestle.

Gratier, M., & Devouche, E. (2011). Imitation and repetition of prosodic contour in vocal interaction at 3 months. *Developmental Psychology, 47*(1), 67–76. https://doi.org/10.1037/a0020722

Gratier, M., Devouche, E., Guellai, B., Infanti, R., Yilmaz, E., & Parlato-Oliveira, E. (2015). Early development of turn-taking in vocal interaction between mothers and infants. *Frontiers in Psychology, 6*(1167), 10–3389. https://doi.org/10.3389/fpsyg.2015.01167

Henning, A., & Zmyj, N. (2016). Infant do, infant see: The role of feedback to infant behavior for the understanding of self and others. *Infant and Child Development, 25*(3), 233–239. https://doi.org/10.1002/icd.1987

Hobson, R. P. (1993). *Autism and the development of mind*. Erlbaum.

Jaffe, J., Beebe, B., Feldstein, S., Crown, C. L., Jasnow, M. D., Rochat, P., & Stern, D. N. (2001). Rhythms of dialogue in infancy: Coordinated timing in development. *Monographs of the Society for Research in Child Development, 66*(2), i–149. https://www.jstor.org/stable/3181589

Jones, S. (2006, June). Infants learn to imitate by being imitated [Paper presentation]. In *Proceedings of the International Conference on Development and Learning: The Tenth International Conference on Development and Learning*, Bloomington, IN, United States. https://www.researchgate.net/profile/Susan-Jones-20/publication/265990133_Infants_Learn_to_Imitate_by_Being_Imitated/links/54ac7fac0cf2479c2ee7eab6/Infants-Learn-to-Imitate-by-Being-Imitated.pdf

Jones, S., & Yoshida, H. (2011). Imitation in infancy and the acquisition of body knowledge. In V. Slaughter & C. A. Brownell (Eds.), *Early development of body representations* (pp. 207–225). Cambridge University Press.

Jonsson, C. O., & Clinton, D. (2006). What do mothers attune to during interactions with their infants? *Infant and Child Development, 15*, 387–402. https://doi.org/10.1002/icd.466

Jonsson, C. O., Clinton, D., Fahrman, M., Mazzaglia, G., Novak, S., & Sörhus, K. (2001). How do mothers signal shared feeling-states to their infants? An investigation of affect attunement and imitation during the first year of life. *Scandinavian Journal of Psychology, 42*(4), 377–381. https://doi.org/10.1111/1467-9450.00249

Kärtner, J., Keller, H., Lamm, B., Abels, M., Yovsi, R. D., Chaudhary, N., & Su, Y. (2008). Similarities and differences in contingency experiences of 3-month-olds across sociocultural contexts. *Infant Behavior and Development, 31*(3), 488–500. https://doi.org/10.1016/j.infbeh.2008.01.001

Kärtner, J., Keller, H., & Yovsi, R. D. (2010). Mother–infant interaction during the first 3 months: The emergence of culture-specific contingency patterns. *Child Development, 81*(2), 540–554. https://doi.org/10.1111/j.1467-8624.2009.01414.x

Kaye, K. (1982). *Mental and social life of babies*. University Chicago Press.

Keller, H., Lohaus, A., Völker, S., Cappenberg, M., & Chasiotis, A. (1999). Temporal contingency as an independent component of parenting behavior. *Child Development, 70*(2), 474–485. https://doi.org/10.1111/1467-8624.00034

Keven, N., & Akins, K. (2017). Neonatal imitation in context: Sensorimotor development in the perinatal period. *Behavioral and Brain Sciences, 40*, E381. https://doi.org/10.1017/S0140525X16000911

Kipp, M. (2008, May). *Spatiotemporal coding in ANVIL* [Paper presentation]. *Sixth International Conference on Language Resources and Evaluation*, Marrakech, Morocco.

Knafo, H. (2016). The development of body image in school-aged girls: A review of the literature from sociocultural, social learning theory, psychoanalytic, and attachment theory perspectives. *The New School Psychology Bulletin, 13*(2), 1–16. http://nspb.net/index.php/nspb/article/view/270

Kokkinaki, T. (2003). A longitudinal, naturalistic and cross-cultural study on emotions in early infant-parent imitative interaction. *British Journal of Developmental Psychology, 21*(2), 243–258. https://doi.org/10.1348/026151003765264066

Kokkinaki, T., & Kugiumutzakis, G. (2000). Basic aspects of vocal imitation in infant-parent interaction during the first 6 months. *Journal of Reproductive and Infant Psychology, 18*(3), 173–187. https://doi.org/10.1080/713683042

Kokkinaki, T., & Vasdekis, V. (2003). A cross-cultural study on early vocal imitative phenomena in different relationships. *Journal of Reproductive and Infant Psychology, 21*(2), 85–101. https://doi.org/10.1080/0264683031000124055

Kokkinaki, T., & Vitalaki, E. (2013). Comparing spontaneous imitation in grandmother-infant and mother-infant interaction: A three generation familial study. *The International Journal of Aging and Human Development, 77*(2), 77–105. https://doi.org/10.2190/AG.77.2.a

Kugiumutzakis, G. (1993). Intersubjective vocal imitation in early mother-infant interaction. In J. Nadel & L. Camaioni (Eds.), *New perspectives in early communicative development* (pp. 23–47). Routledge.

Kugiumutzakis, G., Kokkinaki, T., Makrodimitraki, M., & Vitalaki, E. (2005). Emotions in early mimesis. In J. Nadel & D. Muir (Eds.), *Emotional development* (pp. 161–182). Oxford University Press.

Kykyri, V. L., Karvonen, A., Wahlström, J., Kaartinen, J., Penttonen, M., & Seikkula, J. (2017). Soft prosody and embodied attunement in therapeutic interaction: A multimethod case study of a moment of change. *Journal of Constructivist Psychology, 30*(3), 211–234. https://doi.org/10.1080/10720537.2016.1183538

Malatesta, C. Z., & Haviland, J. M. (1982). Learning display rules: The socialization of emotion expression in infancy. *Child Development, 53*(4), 991–1003. https://doi.org/10.2307/1129139

Malloch, S. & Trevarthen, C. (2008). Musicality: Communicating the vitality and interests of life. En S. Malloch & C. Trevarthen (Eds.), Communicative musicality (pp. 1–11). : Oxford University Press.

Maratos, O. (1998). Neonatal, early and later imitation: Same order phenomena? In F. Simion & G. Butterworth (Eds.), *The development of sensory, motor and cognitive capacities in early infancy: From perception to cognition* (pp. 145–160). Routledge.

Markova, G., & Legerstee, M. (2006). Contingency, imitation, and affect sharing: Foundations of infants' social awareness. *Developmental Psychology, 42*(1), 132–141. https://doi.org/10.1037/0012-1649.42.1.132

Masur, E. F., & Rodemaker, J. E. (1999). Mothers' and infants' spontaneous vocal, verbal, and action imitation during the second year. *Merrill-Palmer Quarterly, 45*(3), 392–412. http://www.jstor.org/stable/23092579

Meltzoff, A. N. (1990). Foundations for developing a concept of self: The role of imitation in relating self to other and the value of social mirroring, social modeling, and self practice in infancy. In D. Cicchetti & M. Beeghly (Eds.), *The self in transition: Infancy to childhood* (pp. 139–164). University of Chicago Press.

Meltzoff, A. N., Murray, L., Simpson, E., Heimann, M., Nagy, E., Nadel, J., Pedersen, E., Brooks, R., Messinger, D., De Pascalis, L., Subiaul, F., Paukner, A., & Ferrari, P. F. (2017). Re-examination of Oostenbroek et al. (2016): Evidence for neonatal imitation of tongue protrusion. *Developmental Science, 21*(4), *e12609*. https://doi.org/10.1111/desc.12609

Moran, G., Krupka, A., Tutton, A., & Symons, D. (1987). Patterns of maternal and infant imitation during play. *Infant Behavior and Development, 10*(4), 477–491. https://doi.org/10.1016/0163-6383(87)90044-0

Nagy, E., & Molnar, P. (2004). Homo imitans or homo provocans? Human imprinting model of neonatal imitation. *Infant Behavior and Development, 27*, 57–63. https://doi.org/10.1016/j.infbeh.2003.06.004

Nomikou, I., Leonardi, G., Rohlfing, K. J., & Rączaszek-Leonardi, J. (2016). Constructing interaction: The development of gaze dynamics. *Infant and Child Development, 25*(3), 277–295. https://doi.org/10.1002/icd.1975

Oostenbroek, J., Suddendorf, T., Nielsen, M., Redshaw, J., Kennedy-Costantini, S., Davis, J., Clark, S., & Slaughter, V. (2016). Comprehensive longitudinal study challenges the existence of neonatal imitation in humans. *Current Biology, 26*(10), 1334–1338. https://doi.org/10.1016/j.cub.2016.03.047

Papoušek, M., & PapouŠek, H. (1989). Forms and functions of vocal matching in interactions between mothers and their precanonical infants. *First Language, 9*(6), 137–157. https://doi.org/10.1177/014272378900900603

Papoušek, H., & Papoušek, M. (2002). Intuitive parenting. In M. H. Bornstein (Ed.), *Handbook of parenting* (*Biology and ecology of parenting*) (Vol. 2, pp. 183–203). Lawrence Erlbaum Associates.

Papoušek, H., Papoušek, M., & Kestermann, G. (2000). Preverbal communication: Emergence of representative symbols. In N. Budwing, I. Užgiris, & J. Wertsch (Eds.), *Communication: An arena of development* (pp. 81–108). Ablex Publishing Corporation.

Pawlby, S. (1977). Imitative interaction. In H. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 203–224). Academic Press.

Pellegrini, A. D. (2009). *The role of play in human development*. Oxford University Press.

Pérez, D. I., & Gomila, A. (2021). *Social cognition and the second person in human interaction*. Routledge. https://doi.org/10.4324/9781003133155

Pratikaki, A., & Kokkinaki, T. (2013). Emotional expressions in grandparent-infant grandchild interaction in the course of the first year of life. *Europe's Journal of Psychology, 9*(3), 531–551. https://doi.org/10.5964/ejop.v9i3.553

Provenzi, L., Scotto di Minico, G., Giusti, L., Guida, E., & Müller, M. (2018). Disentangling the dyadic dance: Theoretical, methodological and outcomes systematic review of mother-infant dyadic processes. *Frontiers in Psychology, 9*, 348. https://doi.org/10.3389/fpsyg.2018.00348

Reddy, V. (2004). Mind knowledge in the first year: Understanding attention and intention. In I. G. Bremner & A. Fogel (Eds.), *Blackwell handbook of infant development* (pp. 241–264). Blackwell publishing. Ltd..

Reddy, V. (2008a). *How infants know minds*. Harvard University Press.

Reddy, V. (2008b). Experiencing others: A second person approach in other awareness. In U. Müller, J. Carpendale, N. Budwig, & B. Sokol (Eds.), *Social life and social knowledge. Toward a process account of development* (pp. 123–144). Lawrence Erlbaun.

Reddy, V., Chisholm, V., Forrester, D., Conforti, M., & Maniatopoulou, D. (2007). Facing the perfect contingency: Interactions with the self at 2 and 3 months. *Infant Behavior and Development, 30*(2), 195–121. https://doi.org/10.1016/j.infbeh.2007.02.009

Rivière, A. (1986/2003). Interacción precoz. Una perspectiva vygotskiana a partir de los esquemas de Piaget [Early interaction. A Vygotskian perspective from Piaget's schemas]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (Eds.) *Ángel Rivière. Obras Escogidas* [Ángel Rivière. Selected Works], Vol II (pp. 109–142). Panamericana.

Rochat, P. (2001). *The infants world [El mundo del bebé (R. Filella, trad.). Morata, 2004]*. Harvard University Press.

Saby, J. N., Marshall, P. T., & Meltzoff, A. N. (2012). Neural correlates of being imitated: An EEG study in preverbal infants. *Social Neuroscience, 7*(6), 650–661. https://doi.org/10.1080/17470919.2012.691429

Sauciuc, G.-A., Zlakowska, J., Persson, T., Lenninger, S., & Alenkaer Madsen, E. (2020). Imitation recognition and its prosocial effects in 6-month old infants. *PLoS One, 15*(5), e0232717. https://doi.org/10.1371/journal.pone.0232717

Schaffer, H. R. (1977). Early interactive development. In H. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 3–16). Academic Press.

Sheets-Johnstone, M. (2015). *The corporeal turn: An interdisciplinary reader*. Andrews UK Limited.

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver* (pp. 187–214). Wiley.

Stern, D. N. (1985). *The interpersonal world of the infant. A view from psychoanalysis and developmental psychology*. Basic Books.

Stern, D. N. (2010). *Forms of vitality: Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford University Press.

Stern, D., Hofer, L., Haft, W., & Dore, J. D. (1985). Affect attunement: The sharing of feeling states between mother and infant by means of inter-modal fluency. In T. M. Field & N. Fox (Eds.), *Social perception in infants* (pp. 249–268). Ablex.

Tomasello, M. (2019). *Becoming human*. Harvard University Press.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny*. Cambridge University Press.

Trevarthen, C. (1999). Musicality and the intrinsic motive pulse: Evidence from human psychobiology and infant communication. *Musicæ Scientiæ*, 155–215. https://doi.org/10.1177/10298649000030S109

Trevarthen, C. (2009). Embodied human intersubjectivity: Imaginative agency, to share meaning. *Cognitive Semiotics, 4*(1), 6–56. https://doi.org/10.1515/cogsem.2009.4.1.6

Trevarthen, C. (2017). Play with infants: The impulse for human story-telling. In T. Bruce, P. Hakkarainen, & M. Bredikyte (Eds.), *The Routledge international handbook of play in early childhood* (pp. 198–216). Routledge.

Trondalen, G., & Skårderud, F. (2007). Playing with affects: ... And the importance of "affect attunement". *Nordic Journal of Music Therapy, 16*(2), 100–111. https://doi.org/10.1080/08098130709478180

Užgiris, I. C. (1984). Imitation in infancy: Its interpersonal aspects. In M. Perlmutler (Ed.), *Minnesota symposium on child psychology* (Vol. 17, pp. 1–32). Lawrence Erlbaum Associates.

Užgiris, I. C. (1999). Imitation as activity: Its developmental aspects. In J. Nadel & G. Butterworth (Eds.), *Imitation in infancy* (pp. 186–206). Cambridge University Press.

Užgiris, I. C., Benson, J. B., Kruper, J. C., & Vasek, M. E. (1989). Contextual influences on imitative interactions between mothers and infants. In J. J. Lockman & N. L. Hazen (Eds.), *Action in social context: Perspectives on early development* (pp. 103–127). Plenum Press.

Van Egeren, L. A., Barratt, M. S., & Roach, M. A. (2001). Mother–infant responsiveness: Timing, mutual regulation, and interactional context. *Developmental Psychology, 37*(5), 684–697. https://doi.org/10.1037/0012-1649.37.5.684

Vincini, S., Jhang, Y., Buder, E. H., & Gallagher, S. (2017). Neonatal imitation: Theory, experimental design, and significance for the field of social cognition. *Frontiers in Psychology, 8*, 1323. https://doi.org/10.3389/fpsyg.2017.01323

Volgsten, U. (2012). The roots of music: Emotional expression, dialogue and affect attunement in the psychogenesis of music. *Musicae Scientiae, 16*(2), 200–216. https://doi.org/10.1177/1029864912440778

Volgsten, U., & Pripp, O. (2016). Music, memory, and affect attunement: Connecting kurdish diaspora in Stockholm. *Culture Unbound, 8*(2), 144–164. https://doi.org/10.3384/cu.2000.1525.1608144

Whaley, K. K. (1990). The emergence of social play in infancy: A proposed developmental sequence of infant-adult social play. *Early Childhood Research Quarterly, 5*, 347–358. https://doi.org/10.1016/0885-2006(90)90026-W

Silvia Español holds a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). She is currently a researcher at the Latin America Institute of Social Research (IICSAL), Faculty of Latin America Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET, Argentina). She coordinates the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc. in Cognitive Psychology, and Director of FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is a certified Somatic Educator in the Feldenkrais Method. Her work deals with cognitive developmental psychology, psychology of music, and the study of human movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/silvia.espanol

Mariana Bordoni has a PhD. in Psychology from Universidad Nacional de Córdoba (Argentina). She participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc. in Cognitive Psychology at Universidad de Buenos Aires (UBA) and FLACSO. She also teaches in FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is the creator and coordinator of the Baby Social Club. She has collaborated in several research projects funded by UBACyT and ANCPyT. Her publications are available at: https://www.aacademica.org/mariana.bordoni

Mauricio Martínez has a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). He participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning at the Faculty of Latin America Social Sciences (FLACSO). He is Professor of Developmental Psychology at Universidad Abierta Interamericana (UAI) and Faculty member of the MsC. in Cognitive Psychology (FLACSO). He also teaches in the post-graduate virtual course The baby's life: psychological development, scaffolding and nurturing environmen" at FLACSO. His research focuses on early perceptual and intersubjective development, childhood play and early intervention in autism spectrum disorders. He is author of Intervención psicoeducativa para niños con trastorno del espectro autista [Psychoeducational intervention for children with autism spectrum disorders] (2015). His publications are available at: https://www.researchgate.net/profile/Mauricio-Martinez-2

Santiago García Cernaz is a psychologist from Universidad Nacional del Comahue (UNCo) and a PhD. candidate at the Universidad Nacional de La Plata (UNLP, Argentina). He is fellow of the National Scientific and Tecnological Research Council (CONICET) at the Laboratory for the Study of Musical Experience (LEEM) in the Faculty of Arts at the UNLP. He is a Faculty member of the UNLP and also a member of the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). He works as form tutor in FLACSO's virtual post-graduate course The baby's life: psychological development, scaffolding and nurturing environment. His line of work focuses on child socio-cognitive development, particularly play and communicative musicality, in the border area between developmental psychology and the psychology of music.

Viviana Riascos is MSc candidate in Cognitive Psychology and Learning at FLACSO (Argentina). Psychologist, graduated from Universidad Icesi, Colombia. She is a Faculty member at Universidad Abierta Interamericana (UAI). She is member of the Research Group of Socio-Cognitive Processes in Infancy and Early Childhood at the Psychology of Knowledge and Learning Department (FLACSO, Argentina). She has participated in several research proyects in the field of educational psychology and developmental psychology with a focus on embodied cognition. Her research focuses on early communication skills from a multimodal perspective.

Chapter 5
Rhythm as a Prototypical Example of the Relationship Between Intersubjective and Perceptual Development During the First Year of Life

Mauricio Martínez and José Manuel Igoa

Abstract The chapter proposes rhythm as a paradigmatic example of the relationship between intersubjective development and the development of multisensory perception during the first year of life. From the perspective of embodied cognition and the second-person perspective, perception is assigned a leading role, highlighting it as the psychological process that makes social exchange possible. However, research conducted under such perspectives does not usually specify which particular perceptual abilities might be involved and which aspects of people's behavior are directly perceived. Against this background, the chapter establishes some reasonable relationships between the development of multisensory perception of rhythm and the intersubjective development of the infant. To this end, first, it assumes that the infant's intersubjective development involves changes in the infant's behavior that go beyond those related to the transition from primary to secondary intersubjectivity. Second, it exposes the intrinsic value of rhythm for the infant's intersubjective development. Third, it describes the ontogenesis of the detection, discrimination, and recognition of multimodal rhythmic patterns. Finally, through a series of examples, it establishes reasonable relationships between some changes in the infant's social behavior and the development of the three multisensory rhythm perception skills.

M. Martínez (✉)
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina
e-mail: mmartinez@flacso.org.ar

J. M. Igoa
Department of Basic Psychology, Faculty of Psychology, Universidad Autónoma de Madrid, Madrid, Spain

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_5

Keywords Rhythm · Direct social perception · Intersubjectivity · Temporal information · Perceptual skills · Multisensory perception

5.1 Introduction

Rhythm is a fundamental quality of experience and the core component of the temporal organization of the stimuli we perceive and the actions we perform. Therefore, it is not only a sensory property of the stimulation we receive but also an inherent element of the actions we perform. Every sequence of events unfolding in time has a rhythmic structure, whether regular (repetitive) or irregular. Rhythm is a complex phenomenon. Its most outstanding and defining characteristic, as well as its function, is the grouping of events in time. This characteristic depends on two subordinate features, which are the relative duration of the events and the accent or prominence of some over others (Nespor & Vogel, 1986). All this configures a metric structure that determines the mode of temporal organization of the succession of events to which it is applied (Jackendoff, 2002).

Rhythm is an intrinsic property of the temporal organization of speech, being, together with intonation, one of the distinctive features of linguistic prosody (Nooteboom, 1997). It is also an essential feature of the organization of musical pieces and their unfolding in time, both at local and global levels (de la Paz Jacquier et al., 2013). Thus, both domains, speech and music, are characterized by the hierarchical organization of sound units (syllables, words, and intonational phrases in speech, and notes, chords, motifs, and phrases in music) in groups of increasing complexity (Fabb & Halle, 2011). The importance given to rhythm in the domains of speech and music should not obscure the fact that rhythm is a phenomenon that we find everywhere in nature, in the physiology and motor activities of humans and other animals, nor that, in our experience, rhythm is not only a form of stimulus, a perceptual object, but also a product of action.

In the present chapter, we are interested in the genesis and development of sensitivity to rhythm in infants during the first year of life. Our main point, already foreshadowed in the title, is that the developmental study of sensitivity to rhythm will help understand and illustrate the close relationship between the development of perceptual skills and of the evolving experiences and capacities of intersubjective contact in infants throughout the first year of life, a relationship of mutual interdependence better described as *dialectic*, and thus different from the mere correlation of factors that is observed during development or a hypothetical linear causal relationship between them. Our contribution is an attempt to justify this dialectical relationship from a conceptual point of view based on the empirical evidence available so far in developmental studies. In this regard, we argue that the perceptual and the social courses of development ought to be explored together, in order to achieve a cross-fertilization between both domains of study. This dialectical perspective, however, is not the dominant trend in contemporary developmental psychology, which remains anchored to the debate between those who advocate a modular and

largely innate or preformed organization of cognitive competencies in ontogenesis (van der Lely, 2005; Carey, 2011; Baillargeon & Carey, 2012) and those who advocate a globalist conception of development, driven by domain-general learning mechanisms (Elman et al., 1996; Lany & Saffran, 2013) (for a critique of this dichotomy, see Karmiloff-Smith, 2015).

Starting from the dialectical and integrated approach to perceptual and intersubjective development that we adopt in this chapter, we will begin with a brief historical review of the different notions of *infant* found in developmental psychological theories from the end of the nineteenth century to the present day. In this review, we highlight the change in perspective from the first conceptions of the prelinguistic baby as a solipsistic subject, immature and disconnected from his/her environment, to a more integrative and complex view of the baby, who is thought to be endowed with very early perceptual and social skills that radically modify the traditional conception. A turning point in this conceptual change is the discovery of very early perceptual and social skills in the infant that crystallize in the concept of *intersubjectivity* (Trevarthen, 1998). However, the theoretical background underlying this new approach keeps a nativist view of the infant's abilities and a conception of development as a transit through discrete stages and evolutionary leaps that depend, for the most part, on the biological maturation of cognitive structures preformed in the infant's mind. From this follows, moreover, the idea that these cognitive structures follow independent lines of development that can (or even should) be investigated separately.

The main problem of the modular and maturationist conception of development, as we argue in this chapter, lies in the fact that it does not provide an explanation of the genesis of the infant's social behavior, but takes it for granted. This is shown in its approach to the investigation of intersubjective behaviors, which are seen from this perspective as pertaining only to the subject of observation, either as the protagonist of the experience (the so-called first-person perspective) or as the object of analysis (the third-person perspective). In contrast to these two one-sided perspectives, we adhere to a second-person perspective (Pérez & Gomila, 2022; Pauen, 2012), which not only pays attention to the infant's reactions to stimuli from the social environment, but also to the adult's behavior, which is complex insofar as it is multimodal, that is, it comprises and integrates stimulation from various sensory modalities. From this perspective, the emphasis should be placed on the interaction between the infant and the adult, not on the individual and isolated behaviors of the participants, and in particular on complex multimodal actions and stimuli. Rhythm is a paradigmatic example of this action-perception binomial and commonly requires, on the part of the perceiver, the ability to coordinate and integrate information from several sensory systems (vision and hearing, at least), that is, to establish intersensory relationships based on this property of stimulation. Rhythm is thus a quality of experience present in the intersubjective exchanges between the infant and its caregivers and is therefore a topic of study that provides the opportunity to examine the evolution of intersubjective contacts at different moments of early development.

After the historical overview of developmental theories presented in Sect. 5.2, we address the concept of intersubjectivity and its evolution during the first year in Sect. 5.3, emphasizing its gradual evolution, as witness to the kind of changes observed in the infant's social behavior during this period. The description of these changes makes it clear that the transition from an early or primary intersubjectivity, characteristic of the first months, to a secondary intersubjectivity, which for some authors takes place around the ninth month of life, does not occur suddenly and abruptly, as some authors maintain, but rather before and after this chronological boundary there are gradual and significant changes in the infant's social behavior. The aforementioned shift from first- and third-person to second-person perspectives is elaborated in Sect. 5.4, taking as a reference the emergence of mentalistic attributional – or theory of mind – capacities and picking up the idea that these capacities evolve progressively, as revealed by the available empirical evidence.

The setting in which the relationship between the early development of perceptual skills and the evolution of intersubjective contact abilities is revealed is the episodes and experiences of social encounters between the infant and his or her caregiver. The ways in which these episodes develop and the forms that the adult's behavior takes in such encounters are addressed in Sect. 5.5. In this analysis, we highlight the dependence of intersubjective contact on the processes of perception and action and on the capacity to perceptually integrate the multimodal character of the stimulation received from the adult. In other words, the infant's confrontation with this complex information requires the establishment of intersensory relations that enable him/her to recognize and respond appropriately to amodal information (not associated with a specific sensory modality) unfolding over time. This gives way to an analysis (see Sect. 5.5) of the rhythmic component in the behaviors of the adult in interaction with the infant, an analysis that highlights the 'scaffolding' role played by the adult in the acquisition of the infant's perceptual skills.

The last sections of the chapter (Sects. 5.6 and 5.7) are devoted, on the one hand, to presenting the skills that the infant progressively develops on a scale from less to more complex in the domain of multisensory rhythm perception (Sect. 5.6) and, on the other hand, to justifying the dialectical relationship between the development of rhythm perception and intersubjective development. To this end, we review a significant number of studies that parallel the findings in both domains and suggest a co-evolution between perceptual and social skills. With regard to the development of the skills referred to in Sect. 5.6, it is worth noting a distinction between the skills of *detection*, *discrimination*, and *recognition* of intersensory relations in the face of sequences of stimuli (visual and auditory) that are displayed in a synchronized or desynchronized manner. Each of these skills is more complex than the preceding one in the evolutionary and chronological scales so that the acquisition of each of them presupposes the mastery of the previous one. The review of the studies offered here follows a chronological criterion, but it is not intended to be complete, as it seems to us more appropriate to justify the link between the perceptual and social developmental domains, often ignored in the scientific literature, than to provide an extensive list of empirical findings on these issues. The chapter concludes with a general discussion of the issues addressed in it.

5.2 Early Intersubjective Development

With few exceptions, psychological theories toward the end of the nineteenth and beginning of the twentieth centuries considered the infant as a solipsistic subject. The infant was seen as a subject initially disconnected from others and from the world (Rochat, 2001), immersed in a confused, blooming, and ruminative experience (James, 1890). He/she was considered an asocial being incapable of orienting him/herself toward others, too immature even to distinguish her/his mother among different people (Schaffer, 1971). More or less explicitly, various ideas about psychological development flourished under this conception. From this perspective, a number of explanations about the development of the infant's social behavior were proposed in particular, such as social learning theory (Miller & Dollard, 1941; Sears, 1951), psychoanalytic theory (Freud, 1940), and cognitive theories of social development (Bower, 1979). All these formulations assumed the so-called secondary socialization drive hypothesis (Bowlby, 1982; Schaffer, 1971). According to this hypothesis, the infant's interest in other members of the species and the desire to be in their company are the result of being nurtured by them. It is assumed that the infant experiences a series of primary somatic drives that arise from its physical constitution (linked to hunger, thirst, or bodily well-being) and that the satisfaction of these drives by caregivers is the origin of the social bond.

From around the 1960s onward, a series of works began to be developed that radically changed the traditional view that was dominant in psychology by that time on the social life of the infant. The forces behind this change of view regarding infant development were the failure to prove Schaffer's (1971) secondary socialization drive hypothesis, together with the observation (drawn from micro-analytic studies) of the interactions between the infant and his/her caregivers (a systematic review of which can be found in Carretero & Español, 2016), and the pioneering work of Fantz (1958) on the perception of infants. From then on, research on early perceptual abilities and the intersubjective experience of the infant have not ceased to provide fresh empirical data, new hypotheses, and very suggestive theories. It was in this context that the work by Colwyn Trevarthen, who is often credited with coining the terms *primary intersubjectivity* and *secondary intersubjectivity*, emerged (Bråten, 1998). Trevarthen's work offered an interesting platform which, as a kind of exoskeleton and beyond the criticisms it has raised, allows to organize the research on intersubjective development during the first year of life. Broadly speaking, the scheme is made up of the two modes in which the infant shares the intersubjective experience with his/her caregivers (i.e., primary and secondary intersubjectivity) and the transition from one mode to the other (i.e., intersubjective development).

The concept of primary intersubjectivity alludes to the expressive exchanges (by means of vocalizations, body movements, facial gestures) that occur between the infant and his/her caregivers during the first months of life, whose coordination is temporally regulated (Trevarthen, 1998). According to Trevarthen (1982), these exchanges reflect a mode of interpersonal cooperation that supports the

development of later social skills. This mode of intersubjective contact begins to develop around the second month of life and extends until the ninth month. The main feature of this mode of exchange is its dyadic character (face-to-face, person-to-person, body-to-body), where the focus of attention of both participants are the interaction itself, the expressive manifestations that one directs to the other, and the mutual adjustment between their emotional states (Trevarthen, 1998).

Throughout his extensive work, Trevarthen (1974, 1980, 1998, 2001, 2015) emphasized, time and again, the fundamental role that the description of so-called protoconversations had for the development of his work. Bateson (1979) defines protoconversations as sequences in which the adult and the infant establish a pattern of alternating vocalizations, without overlapping or interruptions, where mothers usually utter short phrases and infants respond with audible sounds or soft murmurs. It should be noted that the emphasis in Bateson's work was on the alternation of sound exchanges. Based on his own work, Trevarthen extended the concept of protoconversation beyond the vocal expressive pattern to include exchanges that occur on the basis of alternating facial and gestural expressive patterns. A detailed consideration of this extension is provided by Trevarthen (1998). Around the same time, Stern (1971) identified in the interaction of a mother with her 3-month-old baby a temporal pattern based on their movements. This study was able to demonstrate that the infant moved synchronously with the mother only when they were face to face and making eye contact; when this condition was not met, the infant's movements were independent of those of the mother (Stern, 1977).

Thus, the phenomenon of face-to-face dyadic exchange (through the coordination of vocal, facial, gestural expressions, and body movements) between infant and adult for the sole purpose of shared experience was established as the canon of primary intersubjective experience. Stern (1977) referred to this experience in terms of early social play, "a human happening, conducted solely with interpersonal 'moves', with no other end in mind than to be with and enjoy someone else" (1983, p. 91). These events are considered playful because they simply occur in a limited space of time during which one or both members focus their attention on the social behavior of the other participant by also reacting with social behaviors by means of interpersonal movements or actions and without using toys, artifacts, or rules of play, at least during the first six months of life (Stern, 1977).

From the first studies conducted in the early 1970s to date, research on early intersubjective experiences has been an active research topic. Likewise, in the field of developmental psychology, these early social experiences are often attributed a crucial role in the psychological development of the infant (Dissanayake, 2014).

The difference between primary and secondary intersubjectivity is determined by the interaction of the various components that take part in the modes of intersubjective contact: the self, the other, and the object (Trevarthen, 1998). While primary intersubjectivity refers to the coordination of the self with the other based on the correspondence of form, attunement, and intensity of emotional and behavioral expressions, secondary intersubjectivity involves an object in the interaction (Trevarthen, 1998). It refers to the inter-coordination of the self with the other and the object, based on the cooperative exchange of gestures and referential affective

expressions. This mode of intersubjective contact emerges at around 9 months and shows up when the child combines in the same sequence or communicative episode acts of joint praxis with interpersonal acts (Hubley & Trevarthen, 1979). Acts of joint praxis are defined as actions performed on objects that are oriented to capture the attention or encourage the action of the other person, such as pointing, giving, showing, offering, taking, touching with the object, and following instructions. Interpersonal acts are defined as communicative acts directed to adults that do not refer to objects, such as smiling, vocalizing, looking at the other's face, and extending arms toward the other.

During the period in which the secondary intersubjective contact mode emerges, it is possible to identify three particular phenomena that became relevant topics of research for developmental psychology: joint visual attention or deictic gaze (Bruner, 1983; Collis & Schaffer, 1975; Emery, 2000; Gray, 1978), behavioral regulation and joint attention (Mundy, 2016), and the development of pre-verbal gestural communication (Bates et al., 1975; Bates, 1976; Sarriá & Rivière, 1991). These terms are often considered synonyms or overlapping phenomena. However, the distinction between them allows us to identify the arbitrary way in which they allude to very molecular aspects of the experiences of secondary intersubjectivity as they are approached from different theoretical traditions.

5.3 Beyond the Transit

The explanation proposed, and still upheld today, by Trevarthen to account for the transition from primary to secondary intersubjectivity is one of the most controversial aspects of his work: first, because he endorses a nativist account of the emergence of communication between the infant and the adult, and second, because he attributes the change from primary to secondary intersubjective experience to the maturation of the brain mechanisms responsible for intentional behavior. In his initial formulations (Trevarthen, 1977), based on Gibson's idea that perception depends on the actions that the organism performs to extract information from the environment, he postulated that the actions of the newborn can be conceived of as rudimentary intentions insofar as they are oriented to attain a particular kind of experience. During the first months of life, such actions depend on the configuration of the infant's nervous system, so he understands that the separate actions that the infant performs in front of people and objects are innately predesigned. Thus, he interpreted his observations of the infant's behaviors as signs of intentional communication. Likewise, he attributed the continuous increase of the infant's adaptive abilities to the growth of preconfigured structures in the brain (Trevarthen, 1977). Forty years after his initial formulation, these assumptions remain unchanged (Trevarthen, 2015, 2017). In our view, embracing this perspective has important negative consequences for developmental psychology. Accordingly, Rivière (1986/2003), from a constructivist-interactionist perspective, emphasized that by postulating from the moment of birth capacities to communicate, cooperate, and interact, the major aim

of genetic explanations is lost, namely to account for how these capacities develop during ontogenesis.

Nevertheless, the description of the two modes of sharing intersubjective experience proposed by Trevarthen had an important impact on developmental psychology: in addition to offering a description of qualitatively different modes of social involvement, it defined a problem space whose approach generated data and evidence that we deem relevant to understand the experiences of intersubjectivity, as well as the relationship between intersubjective and perceptual development.

The importance of these changes for developmental psychology led some authors to refer to them as true revolutions. For example, Tomasello (2008) refers to the emergence of secondary intersubjectivity experiences as the revolution of the ninth month. Similarly, Rochat (2001) refers to primary and secondary intersubjectivity by labeling them as the second- and ninth-month revolutions. However, research on early intersubjective development (including that conducted by Trevarthen) shows that the infant's increasing and subtle involvement during intersubjective encounters goes beyond the coordination of his/her expressive acts with those of the adult during dyadic interactions, and of his/her praxic and communicative acts during triadic interactions, and that the genesis of the infant's intersubjective engagement comes about under the scaffold of parenting practices and not by the unfolding of an innately pre-established plan (Kaye, 1982).

First of all, it is important to note that the transition from one mode of intersubjectivity to another should not be interpreted in terms of the sudden emergence of a new set of skills in the infant. From his early studies, Trevarthen himself (Trevarthen & Hubley, 1978) documented the progressive emergence of play in which the infant engages jointly with toys and with his/her mother (he described, for example, how a 25-week-old baby actively behaved while her mother animated toys in front of her). From 6 months, the separate intentions that the infant shows toward people and objects undergo systematic integrations leading to a joint interest in purposes toward objects and events during the shared episode, the climax of which occurs at around 9–12 months, a period in which the infant is fully able to coordinate acts of joint praxis with interpersonal acts (Hubley & Trevarthen, 1979). Consequently, the transition from one mode to another of intersubjectivity should be understood as a gradual process.

Second, it is also necessary to take into account that, in addition to the two major achievements that occur from the second and ninth months, there are other relevant changes in the infant's social behavior during the first year of life. From the first month and a half, the baby's expressive behaviors (smiles, vocalizations, movements), which initially manifest themselves in a brief and isolated manner, are functionally reorganized, giving rise to more complex social responses (Kaye, 1982). Between 2 and 6 months, they go from only visually attending to ostensive rhythmic gestures with objects enacted by the adult to performing the first canonical rhythmic gestures with objects themselves (Moreno-Núñez et al., 2015). From the second to the eighth month, infants express in various ways (by combining vocalizations, facial expressions, and gestures) their intention to re-engage with the adult when the adult breaks the dynamics of the interaction (Melinder et al., 2010). Around 6

months of age, the perception and anticipation of social contingencies increases (Riviére, 1986/2003; Kunichan, 2014[1]). From this age, they become active provocateurs with the intention of attracting the adult's attention to the interaction (Reddy, 2008). In addition, they respond in a differentiated manner in front of an adult who imitates or affectively attunes with them (Bordoni et al., 2016). At around 9 months, they segment adult behavior based on rhythmic/dynamic modeling of adult performance (Español, 2004). At 9 months, they also shake objects performing multimodal rhythmic movements. These movements (performed in intersubjective contexts) may serve as an opportunity for the child to learn to orient the attention of another person and are also positively related to proximal deictic gestures performed at 12 months (Murillo et al., 2021). In summary, accounting for intersubjective development implies explaining much more than the unique emergence of primary intersubjectivity experiences, around 2 months, and of secondary intersubjectivity, around 9 months.

5.4 The Psychological Explanation of Intersubjective Development

We have already mentioned that Trevarthen's explanation of intersubjective development is the most criticized aspect of his contributions. The nativist and maturationist claims of his explanation only sever it from the main objective of developmental psychology: to account for its genesis. To account for the genesis of intersubjective development, psychology must identify what processes lie behind the infant's social behavior and explore how changes in these processes are linked to changes in the infant's social behavior. However, there is no general agreement among researchers as to the type of process that supports intersubjective contact. Ultimately, the different psychological explanations available depend on how intersubjectivity is conceived. Gómez (1998) distinguished two conceptions under which research in developmental psychology is conducted: one-way intersubjectivity, which manifests itself when one subjectivity takes another subjectivity as its object, and two-way intersubjectivity, which occurs when two subjectivities come into contact. The first conception refers to the capacity to think about what the other has in mind and is guided by the assumption that behavior is governed by internal representations and computational or inferential processes operating on such representations. The second, on the contrary, emphasizes interaction and the coordination of actions and emotions among the participants in a social event and minimizes, or even rejects, the role of representations in the explanation of behavior.

[1] Kunichan, J. (2014). Percepción y anticipación de contingencias sociales en la génesis de la comunicación intencionada [Perception and anticipation of social contingencies in the genesis of intentional communication]. Unpublished master's thesis. Universidad Nacional de Buenos Aires.

Since Premack and Woodruff (1978) coined the term "theory of mind" (ToM), one-way intersubjectivity became the dominant conception in a field of research whose explanatory horizon is the resolution of the false belief task by the age 4.5 years (Wellman et al., 2001). ToM is the psychological process that allows people to interpret, predict, and make sense of the behavior of others by attributing mental states such as beliefs, desires, etc. (Baron-Cohen et al., 2013). ToM became the bull's-eye of research on the psychological processes involved in the development of social interactions. From the mid-1980s onward, different theoretical proposals emerged to explain the functioning and development of ToM, and the following decade witnessed the establishment of the two major perspectives based on one-way intersubjectivity that dominate the field of research: the first-person perspective (henceforth FPP) and the third-person perspective (henceforth TPP) (Carruthers and Smith 1996, and see the monographic issue published by *Mind and Language* 7(1−2)). From the FPP, for example, Harris (1992) proposed simulation as the main substrate that would allow attributing mental states to others. Taking as a starting point one's own mental states, a person imagines the mental states of another and simulates him/herself in a particular situation. On the other hand, TPP proposed different mechanisms or abilities underlying the capacity to infer mental states from observable behavior, for example, a theoretical knowledge (Gopnik & Wellman, 1994), a decoupling module of representations (Leslie, 1987), or a generic metarepresentational capacity (Perner, 1991).

Most researchers who assume a one-way or third-person intersubjectivity made a vow of acceptance of the genetic role of primary intersubjectivity states (Español, 2003). Thus, Baron-Cohen (2005), Meltzoff (2011), and Tomasello (2008) cite and draw on Trevarthen's idea of intersubjectivity as a basis for the development of later intersubjective skills. However, they base their explanations on inferential mechanisms or capacities that, in our opinion, incur in some inconsistencies.

For example, Baron-Cohen (2005) postulated three cognitive sub-mechanisms for his ToM/Empathy modular system that would be active before 9 months: emotion detector, intentionality detector, and gaze direction detector. These three subcomponents (which allow the infant to infer and attribute mental states, e.g., intentions) would explain the infant's experiences of primary intersubjectivity. Beyond the classic criticisms of the model (e.g., the assumed nativism, its failure to explain the emergence of joint attention as a reaction or an initiative at different ages, among others), one might ask which of all the postulated mechanisms is responsible for processing the temporal regularities that model adult behavior and that, ultimately, turn out to be the condition of possibility for the intersubjective encounter between the infant and its parenting figures to take place.

For his part, Meltzoff (2011), Meltzoff and Williamsom (2017) proposes that knowledge of other minds depends on a *like-me* inference through which the infant understands that he/she can act like others and others can act like him/her. The formulation proposed by Meltzoff is based on his studies on neonatal imitation (Meltzoff & Moore, 1977), for which he postulated the existence of a supramodal-abstract code that supports Active Intermodal Mapping (Meltzoff & Moore, 1997), which, in turn, would be biologically supported (Saby et al., 2012) and linked to the

mirror neuron system (Iacoboni, 2008). This would allow the infant to make an inference of the type "I am like you," which means that there is an I and a you to begin with, giving rise to experiences of intersubjectivity of a cognitive nature, since they involve a recognition of others and *like me*, a recognition of the similarity between oneself and others (Español, 2003). This conception was already discussed at the time by Rivière (1986/2003), who also postulated the need to posit a dialectical relationship between intersensory relationship competencies (the proposed supra-modal code) and social relationship competencies. Furthermore, Meltzoff and Moore (1998) mentioned that their explanation of intersubjectivity based on imitation is not opposed to that of Trevarthen, but rather complements it, although they explicitly leave aside the phenomena of temporal synchrony, mutual timing, or temporal tuning, since they doubt that sensitivity to time really requires the attribution of psychological properties to the interaction partner (or, rather, to the partner in mutual imitation games).

In our view, Meltzoff and Moore make two mistakes in their explanation of early imitation skills. First, they are wrong in separating the perceptual component involved in intersensory relatedness skills from the emotional exchange inherent in intersubjective encounters, since early imitation takes place in an intersubjective context and thus cannot be reduced to a face-to-face exchange of facial configurations. It is therefore inseparable from the emotional exchange in which it occurs and is inherently linked to the forms of vitality that shape all human actions. The second mistake, probably derived from the previous one, is to ignore the fundamental role of temporal synchrony in the development of this capacity. Probably due to its controversial nature, the phenomenon of neonatal imitation remains a relevant research topic today (Davis et al., 2021).

Another example of the attribution of inferential capacities in infants during the first year of life is found in Tomasello (1999, 2008), who assumes that joint attention and gestural communication emerge around 9 months because infants develop the capacity to "understand that others are intentional and rational agents like me" (2013, p. 107). Regarding the previous months, Tomasello mentions the affective attunement as described by Stern et al. (1985) and the infant's ability to adjust its emotional states to those of the adult. However, it is not clear how this adjustment occurs, or why different responses to attunement and imitation are given between 6 and 9 months (Bordoni et al., 2016), when intentional behaviors (which require a means-end distinction), that enable to understand that others are intentional and rational agents like oneself are not yet observed in the infant's repertoire. Moreover, the segmentation of behavior, both one's own and that of our interaction partner, is crucial to distinguish means from ends. The question then arises as to how an infant who has not developed an understanding of intentionality could segment adult behavior.

Some authors, underpinned by TPP, argue that the ability to infer mental states is present very early in infants (Gergely & Csibra, 2003; Onishi & Baillargeon, 2005). However, a recent meta-analysis carried out with 33 studies and 56 experimental conditions (Barone et al., 2019) revealed that the results varied significantly depending on several factors, such as the type of paradigm used in the research – i.e., the

expectation violation paradigm yielded a better performance than the anticipatory looking or interactive paradigms – or even the date of publication of the study – with more recent studies showing a poorer performance of the infants.

Although the mainstream approach in the studies on the development of social cognition continues to be the one-way intersubjectivity and the two perspectives derived from it, two-way intersubjectivity is gaining more favor in early intersubjective development research. Interestingly, however, until the emergence of ToM as a catalyst construct in research on the development of social cognition, two-way intersubjectivity was the dominant conception in studies of infant social development (Schaffer, 1977; Kaye, 1982). On the other hand, since the beginning of this century, an alternative perspective to the first- and third-person perspectives began to emerge: the second-person perspective (hereafter SPP) (Pérez & Gomila, 2022). The work developed by authors such as Bruner, Bateson, Trevarthen, Stern, Kaye, Schaffer, or Fogel, among others, approached the study of the early intersubjective experience of the infant by emphasizing the interactive and temporal dimensions of the social encounter by means of microanalysis techniques (Schaffer, 1977). In this regard, Kaye (1982) remarked that parents and infants constitute a *social system*, establishing from very early on a shared understanding between individuals.

A relevant issue that we wish to highlight in this approach to the study of intersubjective development is the recurrent reference to the perceptual capacities that enable social encounters to happen between infants and their caregivers. For example, Schaffer (1971) emphasized that the infant's sociability has its origin in perceptual encounters with other people. Similarly, Hobson (1993) – one of the first researchers contemporary to ToM to question the need to postulate theoretical-inferential or simulationist capacities for social interaction – made the right choice by putting the focus on perception. He claimed that many of the infantile capacities for social relations are based on some kind of perceptual foundation, concluding that for babies to be capable of participating in certain kinds of interpersonal exchanges at 2 months of age, they must have a sufficiently adjusted perceptual system (Hobson, 1993).

SPP emerged, in clear consonance with the criticisms made of classical cognitivism by embodied psychology (Varela et al., 1991) or the 4E approach (corresponding to the properties of cognition termed Embedded, Embodied, Extended, and Enacted) (Rowlands, 2010), as an alternative approach to the first- and third-person perspectives linked to one-way intersubjectivity. Under this perspective, it is assumed that we experience others' minds directly and immediately, without the need for theories, inferences, or simulated projections; we perceive others' minds in their expressions as we engage in an interaction (Pérez & Gomila, 2022). Understanding others from FPP and TPP implies detecting (inferring or imagining) that behind the behavior, at the level of the unobservable mental, there are phenomena (beliefs, desires, etc.) that cause the observed behavior. On the other hand, from the SPP perspective, understanding others' behaviors involves perceiving that there are structures or forms (*Gestalten*) in those behaviors that link them directly with other people (Gómez, 2008, 2009). This link does not belong to an unobservable level (as it happens with desires or beliefs); on the contrary, it belongs to the

same – observable – level as the behavior itself (Gómez, 2010). Pérez and Gomila (2018) listed a series of characteristics that define the SPP:

1. The understanding of people occurs only when they experience dynamic interactions face to face or, better said, body to body. In this sense, it is an *online* understanding where the physical presence of the other is necessary.
2. The expressive aspects of the body are seen as expressive (not interpreted) and therefore understanding cannot occur unless the body is present (point 1).
3. Understanding is reciprocal. During the interaction, the understanding of the behavior of one of the participants inevitably modifies the behavior of the other.
4. This mode of understanding is not founded on propositional attitudes (desires or beliefs), but on dynamic affects and emotions.
5. Second-person interactions are not *goal-oriented* – i.e., intended to achieve a mental state about the mental state of the other – but induce a mental state caused by the mental-bodily-expressed state of the other and so on.
6. This process is linked to action insofar as it involves public, overt bodily action, and emotional behavior, i.e., facial alterations and other behaviors such as flee or consoling behaviors.
7. The existence of a world of shared references is not mandatory.
8. It does not require language.
9. The kind of understanding stemming from the SPP is phylogenetically, ontogenetically, and logically prior to the kind of understanding entertained by the FPP and the TPP.
10. This mode of understanding, being prior, as defined in 9, is necessary to acquire the kind of mental concepts involved in the prototypical attributions of FPP and TPP.

It is worth clarifying at this point that the embodied cognition or 4E approaches do not constitute a monolithic and homogeneous block but comprise a variety of theories that differ in some of their postulates, from the most radical ones that categorically reject the notion of representation as a construct with explanatory value, to other more pluralistic ones that admit the existence of minimal representations that may mediate or intervene in some forms of action, such as the so-called "action-oriented representations" (see Rowlands, 2006; Gallagher, 2008).

Regardless of these differences, we want to stress two issues about the SPP. First, some researchers propose to reinterpret from the standpoint of SPP the data garnered (though but not entirely conclusive) from studies conducted to test inferential mentalistic attribution abilities in pre-verbal subjects. For example, Barone and Gomila (2020) argue that infants' performance in indirect false belief tasks can be explained in terms of an elementary form of mental attribution that simply requires that the infant know how to identify the object that is in the focus of the agent's attention and maintain for some time in working memory both the object and its locations, as well as the gazes directed at it. Second, this perspective assigns a very prominent role to perception. Thus, Scotto (2017) states that the understanding of others develops from birth, depends essentially on perception, and manifests itself in practical contexts, so it is a pragmatic capacity that does not presuppose the

capacity to verbalize reasons. Echoing this same idea, Barone and Gomila (2020) argue that face-to-face interaction promotes a form of attribution that allows to recognize, by means of expressive cues, those mental states that exhibit transparent, implicit, and non-propositional contents.

As we have shown, perception stands out as a highly relevant factor both in research that approaches intersubjective development from the analysis of co-constructed interactions between adults and infants and in theoretical proposals that attempt to account for the psychological processes that enable social exchange. Likewise, it is worth stressing that in the field of social cognition, the notion of direct social perception was presented as a mechanism responsible for social interactions (Abramova & Slors, 2015; Bohl & van den Bos, 2012; Gallagher, 2008; Krueger, 2018; Osler, 2016; Spaulding, 2015; Wiltshire et al., 2015). While the organism's perceptual abilities have become a central topic of the research agenda, some issues need further analysis. First, it is necessary to determine the kind of information involved in the perception of others, which has not yet been fully elucidated in social cognition studies, and in particular those devoted to direct perception (Wiltshire et al., 2015). Second, it remains unclear exactly which perceptual processes are linked to social interaction. Finally, at the ontogenetic level, the relationship between the development of perceptual abilities and changes in the infant's intersubjective behavior during the first year of life has not been clearly established. The question of the relationship between perceptual abilities and intersubjective development has been approached in at least three ways: (1) by taking such relationship for granted, (2) by establishing age-dependent correspondences, and (3) by establishing reasonable relationships between them. The remainder of this chapter is intended to present a reasonable link between the development of very specific perceptual skills and very distinctive changes in the infant's intersubjective behavior during the first year of life.

5.5 The Behavior of the Adult During the Intersubjective Encounter

A crucial issue to understand the link between the development of perception and early intersubjective development is the need to specify what the infant perceives during intersubjective encounters. We have already mentioned that from the SPP approach, it is argued that during intersubjective exchanges participants directly perceive the emotions and intentions of interaction partners. However, if we pause to examine the interaction between an adult and a baby, it turns out that the emotions, for example, are not good candidates of what the baby actually perceives. Traditionally, the concept of emotion is directly associated with the manifestation of a set of facial expressions of basic emotions: anger, happiness, sadness, disgust, surprise, and fear (Ekman, 1982). However, Stern (1977) described how adults, when facing infants, express emotions through their facial expressions in a different

way from what they do when facing another adult. In front of the infant, the adult exaggerates the configuration of the elements that make up the facial expression; for example, eyes and mouth open more widely, while facial expressions have longer duration and change at a slower pace. Thus, from the infant's viewpoint, time-bound changes in the configuration of facial expression during the interaction – its slow blooming and gradual fading – are more salient than the configuration itself (Stern, 1985).

Throughout the first year of life, infants share various social experiences with adults: they receive food, hygienic care, and arousal regulation and share a form of early social play. Although adult behavior has been studied across the different social experiences that infants share with adults – for example, Kaye (1982) studied adult behavior during infant feeding – the analysis of adult behavior during early social play is particularly relevant because, as Stern (1977) points out, during these social encounters the focus of the activity is on sharing the intersubjective experience, being in touch with the other is the only underlying motive during the social encounter. Research has identified different ways in which adult behavior is organized: infant-directed improvised performance (see Chap. 2); affective attunement (Stern, 1985, see Chap. 4 in this volume); imitation (Bordoni 2015,[2] see Chap. 4 in this volume); proto-conversations (Bateson, 1979, see Chap. 3 in this volume); and action songs (Eckerdal & Merker, 2009).

The analysis of the different ways in which the behaviors through which adults create conditions for mutuality or intersubjective contact are expressed revealed a common characteristic: multimodality. Adults offer simultaneous information for the infant's different perceptual systems, information that keeps certain relationships that remain invariant during the interaction. For example, I. Martínez (2014) mentions that in various scenes of adult–infant interaction, multimodal behavior presents a sound-kinetic form in which verbal–gestural match frequently occurs. Similarly, from the standpoint of communicative musicality, Malloch and Trevarthen (2009) state that quality refers to the modulated contours of expression that develop over time. These contours consist of the psychoacoustic attributes of vocalizations (timbre, pitch, loudness) and the attributes of movement direction and intensity. Both attributes often coexist multimodally, "so that a wave of the hand will accompany a 'swoop' of the voice" (Malloch & Trevarthen, 2009, p. 4).

If we consider how a body movement can accompany a "swoop" of the voice, we can identify at least two types of multimodal correspondence or coexistence of the sound-kinetic form. On the one hand, there is correspondence or coexistence relative to quality. Most probably, the rise (in the spatial plane) of the hand movement coincides with the rise (in the plane of the pitch of the sound, i.e., a change from a lower to a higher pitch) of the sound of the vocalization. On the other hand, there is correspondence or coexistence relative to the temporal dimension. That is, the

[2] Bordoni, M. (2015). El establecimiento de coincidencias en las interacciones adulto-bebé. Un estudio longitudinal cuasi-experimental sobre imitación y entonamiento afectivo. [Matching establishment in adult-infant interactions. A longitudinal quasi-experimental study on imitation and affective attunement]. Unpublished doctoral dissertation. Universidad Nacional de Córdoba.

duration of the arm movement (the time that elapses between the beginning and the end of the movement) will coincide with the duration of the voice movement (the time that elapses between the beginning and the end of the sound) and will also coincide with the movement of the orofacial area while the adult utters the sound (the time that elapses between the beginning and the end of the lip movement). The multimodal character of adult behavior confronts the infant with two types of intersensory relations between the information he/she receives: those based on amodal temporal information (which can be perceived indistinctly by different perceptual systems: duration, rate, rhythm) and those based on single modality information (that which can only be perceived through a single perceptual system: color, smell, timbre, tone, temperature, etc.).

If we assume, as proposed by the direct social perception hypothesis espoused by embodied cognition, SPP, and studies of adult–infant interactions, that intersubjective contact depends on perception-action processes and that multimodality and amodal temporal information (duration, rate, and rhythm) are crucial components of adult behavior, investigating the development of multisensory perception of amodal temporal information will allow us to deepen our understanding of intersubjective development. The development of different perceptual abilities during the first year of life (see Chap. 6) would allow us to understand the changes in the infant's participation during intersubjective encounters throughout the first year, without the need to appeal to psychological processes responsible for inferences or simulations.

5.5.1 Rhythm in Infant-Directed Adult Behavior

The information provided by the adult is characterized by its complex temporal organization, in which the rhythmic component plays an essential role (Trevarthen, 1998). Rhythm is a key element of human development. At a biological level of analysis, it has been observed that during episodes of interactive synchrony moms and babies coordinate their heart rates (Feldman et al., 2011). Behaviorally, rhythmic movements have been found to be the largest proportion of adult behavior when interacting with infants. Koester et al. (1989) analyzed 3 min of social play between mothers and their 3-month-old infants by identifying rhythmic episodes in tactile (stroking, patting, and tickling), kinetic (moving the infants' arms or legs), vestibular (rocking the infant), and visual (nodding or shaking the head, playful tongue movements) stimulation that adults offered to infants. They found that, on average, adults offered rhythmic episodes during 48.30% of the three minutes of interaction, and that mothers, sensitive to infants' alertness, adjusted their performance in such a way as to redirect and sustain their attention. Parenting figures have a varied, rhythmically modeled behavioral repertoire that is used to modulate the infant's emotional-behavioral state (Field, 1978). They can thus lead him/her to attain the level of attention necessary for information processing; they can calm him/her or help him/her in the transition to sleep (Papoušek & Papoušek, 1981). Rhythm is so significant in early life that it has been suggested that the mutually conscious

empathy that occurs between adult and infant stems from the infant's rhythmic sensitivity (Trevarthen, 1998).

The presence of rhythm in adult behavior was also studied in interactions where adults incorporate objects into the interaction with infants. Moreno-Núñez et al. (2015) analyzed the interaction of three dyads at 2, 4, and 6 months and found that rhythmic sound components accompany the ostensive gestures that adults perform with objects. They thus showed that rhythm is a key element of the semiotic system that enables adult–child communication with an object. Trevarthen and Hubley (1978) described how a 25-week-old girl actively engaged in a game in which her mother animated toys using some of the elements observed during dyadic intersubjective encounters. The mother moved the toys by performing repeating rhythmic cycles, changing the proximity, direction, and orientation of the toy, and performing rattling and tapping.

It has also been observed that, around the middle of the first year of life, rhythmic melodies appeal to infants. They respond to changes in pulse, intensity, pitch, and timbre, and parents use songs, crooning, and humming to entertain, distract, or calm them to regulate their state of arousal (Trevarthen & Reddy, 2007). Parents also make use of rhythm to scaffold the acquisition of complex skills. When the infant is about 5 months old, they often direct repeated syllable sequences characterized by rhythmic sequencing and overlapping melody. This information serves as a model for repetitive canonical babbling involving segmentation of the vocal stream into syllables (Papoušek, 1996). When adults play interactive games with sequences of rhythmic and repeated actions, they provide the appropriate context for the baby, at around 10 months of age, to be able to decompose and organize the adult's action sequence according to its components, to anticipate the action to come and to perform it before it happens. In this way, the baby becomes an increasingly involved and active partner in the game (Español, 2004, 2005). Fogel and DeKoeyer-Laros (2007) identified how the mother's multimodal, rhythmically modeled behavior in early social play enabled the infant's gaze alternation. The mother and daughter began to engage in this game when the child was 6 months old. The game consisted of joint tapping on the board of the feeding chair. The rhythmic organization of the adult behavior progressively modeled the child's tapping, allowing a stable pattern of joint action to be organized by 10 months of age, which served as a framework for the emergence of the alternation of gaze between the mother's tapping hand and her face. Finally, it is worth mentioning that mothers often respond to the infant's behavior by affectively attuning the infant's behaviors, for example, by mirroring the rhythmic pattern of the infant's behavior through another sensory modality (Stern, 1985).

In the following, we will present some examples of rhythmic modeling of the adult's behavior in different ways the adult displays during the intersubjective encounter.

5.5.2 Rhythm in Performances Directed to the Baby

One of the characteristics of the performance directed to the baby is its repetition-variation form (see Chap. 2). In a scene described and analyzed in Español (2014), a mother is sitting on the floor in front of her baby and the baby is lying on a pillow in front of the mother. After a while of interaction, the infant orients visually by interrupting contact with the mother. When the baby moves and moves away just a little, the mother says, "Where are you going." The mother's behavior comes to be structured in a sentence containing the reason (the utterance) and its repetition three times. In the last two repetitions, the mother adds percussive movements of light pressure with both hands on the baby's trunk. The rhythm of the percussive movement is perfectly adjusted (each hand movement is synchronized with the speech segmentation: "where-are-you-going") to the rhythmic variation of the vocalization. The mother multimodally elaborates the speech unit by progressively adding temporal information (the auditory rhythm is supplemented by tactile rhythm felt on the trunk), underscoring the intersensory redundancy that is added in the last two repetitions. In this case, modeling through rhythmic intersensory redundancy is at the service of variation. Each variation is both familiar and new, quite appropriate for the infant to learn to identify the invariants of adult behavior, which parts of a complex behavior can be suppressed and which other parts must be retained in order to remain the same (Español, 2003).

5.5.3 Rhythm in Affective Attunement

Affective attunement is an adult behavior by which the adult matches the forms of vitality of his/her actions with the forms of vitality of the infant's actions in different social exchange situations (Stern, 2010). Affective attunement has been observed in foster figures interacting with their two-month-old infant (Español et al., 2018). In affective attunement, the adult takes some of the infant's behavior and rephrases it, usually in another modality. Stern et al., 1985 defined it as a particular kind of partial imitation by the adult that involves the performance of overt behavior distinct from that of the infant in which certain features of the original infant behavior are maintained and matched. The features that can be matched are the intensity (absolute or intensity profile), the temporal pattern (duration, rhythm, pulses), or the spatial pattern. By means of affective attunement, the adult intends to indicate to the infant that he/she shares the affective state he/she is feeling when performing his/her action (Stern, 1985; Stern et al., 1985), and in other words, that he/she shares his/her vitality affects (Stern, 2010). For the infant to be sensitive to this behavior, he/she must have developed the ability to establish intersensory relationships in order to identify the relationship between the temporal pattern that shapes his/her own behavior and the temporal pattern that shapes the behavior of the adult interacting with him/her. The infant's visual orientation behavior to affective attunement

changes from 6 to 12 months. During the second half of the first year of life, the frequency of glances directed at the adult increases when the adult imitates the infant, and the frequency of glances toward the adult decreases when she affectively attunes the infant (Bordoni et al., 2016). Stern exemplifies this as follows:

> A nine-month-old boy bangs his hand on a soft toy, at first in some anger but gradually with pleasure, exuberance, and humor. He sets up a steady rhythm. Mother falls into his rhythm and says, "Kaaaaaaabarn, kaaaaaaa-barn," the "barn" falling on the stroke and the "kaaaaa" riding with the preparatory upswing and the suspenseful holding of his arm aloft before it falls. (Stern, 1985, p. 176)

5.6 The Development of Multisensory Perception of Rhythm

The development of multisensory perception of rhythm comprises a set of different and increasingly complex skills: detection, discrimination and recognition (Lewkowicz, 2000; Walker-Andrews, 1997). *Detection* only indicates that an observer is sensitive and responsive in some way to information, i.e., when the organism is affected by a certain level or magnitude of information available to a perceptual system. *Discrimination* refers to the ability to differentiate between two or more objects, events, or between two corresponding aspects (i.e., information) of the same object or event. The ability to discriminate presupposes the ability to detect. One cannot discriminate that which is not detected. *Recognition*, which in turn presupposes detection and discrimination, refers to the ability to establish relationships (generally in terms of their equivalence or perceptual invariance) between two, or more, amodal properties of an object or event when they are presented simultaneously. As regards the development of these three skills, we have empirical evidence corresponding to intrauterine life and the first year of life. As it is logical to assume, and as we will see below, the detection, discrimination, and recognition abilities follow one another in an evolutionary sequence throughout this period.

5.6.1 The Detection of Rhythmic Patterns

Provasi et al. (2014) argue that there is empirical evidence to conclude that the fetus detects and differentiates rhythmic vestibular-tactile-somatosensory stimulation. The works reviewed by Provasi et al. (2014) were conducted with near-birth fetuses and preterm infants. Physiological measures (cardiac or respiratory rate) are used to study the perceptual abilities of fetuses and infants, and the mother's movements (rhythmic gait pattern or anterior–posterior and lateral sway) are used as a source of sensory information.

Electroencephalography recorded from newborn infants (during the second or third day after birth) while listening to rhythmically patterned sound sequences accounts for changes that could be interpreted as detection of rhythm violation in

the sound sequence. The sound pattern with downbeat omission elicited brain activity when such omission was produced occasionally within the context of a regular metrical rhythmic sequence (Winkler et al., 2009).

One issue that we deem worth emphasizing about the data recorded with fetuses concerns their theoretical import. Taking into account the temporal structure and hierarchical organization of the stimuli used (Lewkowicz, 2000) in the studies reviewed by Provasi et al. (2014), we believe that the data should be interpreted in a most parsimonious way. Hence, we suggest they should be considered as evidence for the detection of rhythmic patterns in unborn and preterm fetuses.

5.6.2 Discrimination of Rhythmic Patterns During the First Year of Life

Bahrick and her colleagues studied the ability to discriminate simple rhythmic patterns presented unimodally (movement) and bimodally (sound and movement). They found that 5-month-old infants are able to discriminate between two rhythmic patterns presented bimodally, but at the same age, infants cannot discriminate between two patterns presented unimodally or presented bimodally if the stimuli are presented out of sync (Bahrick & Lickliter, 2000). At 7 months, infants can discriminate between two rhythmic patterns when the tempo is preserved (Pickens & Bahrick, 1995), and also when the original tempo of one of them is varied (Pickens & Bahrick, 1997). The latter suggests that infants categorize rhythmic patterns independently of their tempo of performance (Pickens and Bahrick, 1997). At 8 months, infants discriminate rhythmic patterns when they are presented both bimodally and unimodally (Bahrick & Lickliter, 2004).

Lewkowicz (2003) also reported a series of results on the ability to discriminate rhythmic patterns in bimodally presented synchronized and non-synchronized (sound and motion) information. Infants at 4, 6, 8, and 10 months discriminated between two rhythmic patterns coming from an object (hammer tapping a surface) or a person (a woman repeating the syllable/ba/). Only at 10 months did they discriminate between two rhythmic patterns coming from a person (woman repeating the syllable/ba/) when the information was presented non-synchronously. At 4, 6, and 8 months, infants only discriminated between two patterns coming from one person (female repeating the syllable/ba/) whose information was not synchronized and was not rhythmically modeled. Taken together, the results indicate that from 4 months of age infants discriminate according to rhythmic structure, but only from 10 months of age can they discriminate according to rhythmic structure even when there is no synchrony between visual and auditory information. In turn, Lewkowicz and Marcovitch (2006) reported that infants aged 4, 6, 8, and 10 months discriminated between two rhythmic patterns – that were complex according to the authors – presented in a bimodal manner even when the tempo of the patterns was varied.

5.6.3 *Recognition of Intersensory Relationships Based on Rhythm*

To our knowledge, the recognition of rhythm-based intersensory relationships had not been studied previously. So we decided to conduct a study in our laboratory with 4-, 7-, and 10-month-old infants, in order to learn about the development of this perceptual skill (Martínez et al., 2018). Using the intersensory preference technique, we presented infants with two visual stimuli in which the same woman (seen from the front up to the middle of her torso) is observed uttering the syllabic sound/ta/ (producing the/t/sound by touching the tip of the tongue on the ridge just behind the upper front teeth, and with slightly parted lips, and followed by the vowel/a/) while moving her hands (closing and opening them as if pressing an air pump). Her hands are placed at face level and on both sides of the face with the palms facing forward. A schematic representation of the stimuli used can be seen in Fig. 5.1. In each of the two trials in which the infants participated, one of the visual stimuli was matched to the sound corresponding to one of the two rhythmic patterns that modeled the vocalizations and movements that the woman made with her hands.

In order to establish when infants recognized the intersensory relationship, we contrasted the average proportion of total gaze time of each age group against the random 0.50 value. The results indicated that only the 10-month-old infants visually attended longer to the monitor displaying the video where the movement was synchronized with the sound. Thus, we may conclude that recognition of rhythm-based intersensory relationships is a perceptual skill that is accomplished at 10 months and not before.

By way of summary, we would like to mention that the detection, discrimination, and recognition of rhythmic patterns are skills that emerge at different times during a period that extends from the last months of gestation to the last three months of the first year of life. The detection of multimodal rhythmic patterns is a perceptual skill that is present in the fetus during the last months of gestation and in the first months of life. Discrimination of multimodal rhythmic patterns is an ability that emerges at around the fourth month. And the recognition of rhythm-based intersensory relationships is a skill that shows up around the tenth month of life.

Given the major aim of this chapter, in the previous sections we have only described at what ages the three multisensory perception skills linked to rhythm emerge, but we have not attempted to explain their ontogenesis. Chapter 6 offers a theoretical model of the development of these skills in terms of the infant's participation in intersubjective encounters.

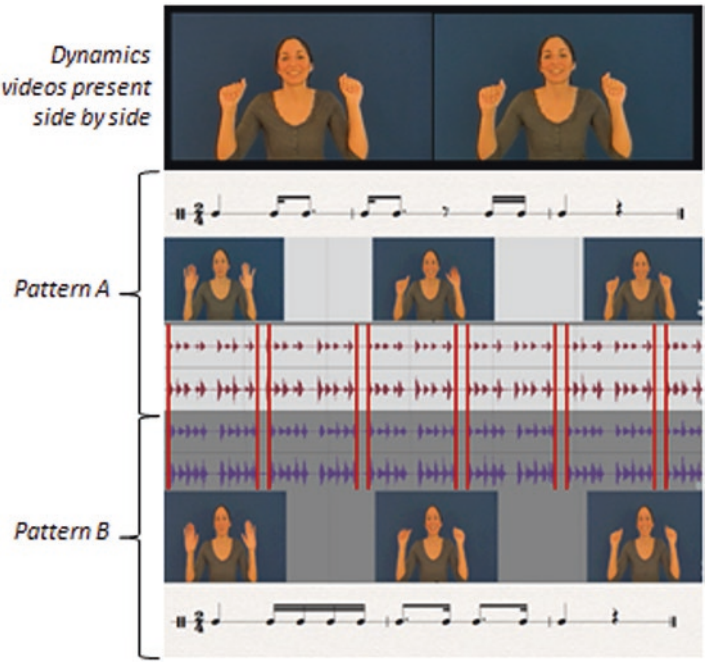


Fig. 5.1 Schematic representation of the two dynamic videos used as stimuli in the experiment. The red lines show the beginning and end of both rhythmic patterns when the two videos are played simultaneously. The musical notation describes the rhythmic pattern performed by the woman in each of the videos. (Reproduced from Martínez et al., 2018)

5.7 Plausible Relations Between the Development of Multisensory Rhythm Perception and Intersubjective Development During the First Year

In the previous sections, we detailed how rhythm shapes the behavior of caregivers as they engage in intersubjective encounters with infants (Sect. 5.5) and the empirical evidence regarding the development of different multisensory perceptual skills linked to rhythm (Sect. 5.6). In this section, we will proceed to suggest plausible relations between the development of different perceptual skills and changes in infant behavior in those intersubjective encounters in which the behavior of their caregivers is rhythmically modeled.

From a dialectical standpoint, according to which the factors involved in a behavioral process or phenomenon do not have a linear causal relationship among them, but instead each one serves as a mediator or stimulus for the others, it is plausible to postulate that, at different ages, the infant's behavior (during those intersubjective encounters where the adult's multimodal behavior is rhythmically modeled) will be different depending on the development of his or her abilities to perceive rhythm. However, to our knowledge, we do not have reliable information to establish

possible links between perceptual development and intersubjective development in order to analyze their possible dialectical relationships as those supported by the SPP. Consequently, the first step to pursue this goal is to identify possible coincidences between the emergence of some of the perceptual skills described in Sect. 6 and changes observed in the infant's social behavior.

5.7.1 The Detection of Multimodal Rhythm Patterns During Gestation and the First Months of Extrauterine Life

Recall that the ability to detect indicates that the organism is sensitive and responds in some way to the information and that this ability is usually found in physiological (changes in heart rate or brain electrical activity) and behavioral (changes in head orientation or gaze direction) responses.

At the physiological level, Lecanuet and Jacquet (2002) reported that toward the end of gestation there are changes in fetal heart rate in response to the absence of movement (resting state) and in response to rhythmic displacement of the mother (in the anterior–posterior direction produced by a rocking chair and in the lateral direction produced by a swinging hammock). During the experiment, mothers moved in two-time sequences (one of 5 s and the other of 26 s) following two rhythmic rocking patterns (one at 2.6 s and the other at 2 s for a full cycle). The average FHR observed during the 5-s or 26-s swing sequence was significantly higher than that obtained by the group whose mothers were at rest. In addition, the average FHR observed during the 26 s of anterior–posterior rocking was significantly higher than the average FHR observed during 26 s or 5 s of lateral rocking.

Feldman et al. (2011) observed face-to-face interactions of mothers with their 3-month-old infants while recording the heart rate of both participants. Microanalysis of the behavior of both participants identified episodes of gaze, affect, and vocal synchrony. During these episodes, the concordance between maternal and infant biological rhythms was significantly increased compared to non-synchronous moments. Mothers adapted their heart rate to that of the infants as the infants adapted their heart rate to that of the mother. This adaptation, which occurred at lags of less than 1 s, resulted in biological synchrony around heart rate acceleration and deceleration.

From a behavioral perspective, Field (1978) proposed that adults modulate the infant's behavior by matching the rhythm of the infant's actions and movements. And similarly, Papoušek and Papoušek (1981) argued that adults display a variety of rhythmic patterns to modulate or regulate the infant's emotional state. The rhythmic and multimodal modeling of adult behavior appears to be in the service of regulating the infant's behavior for different purposes. Because infants demonstrate sensitivity to rhythm during the first months of their lives (Lewkowicz, 2000), it can be interpreted that their response to rhythmic stimulation by adults is largely dependent on this sensitivity.

While adult behavior may increase or decrease the infant's level of alertness, the essential issue for caregivers is rarely about encouraging or arousing the level of attention. Rather, the problem lies in reducing excessively high levels of arousal (Schaffer, 1971). The most commonly used strategy is to hold the infant and provide stimulation by rocking the adult's body (Korner & Thoman, 1970) together with vocalizations and stroking or patting the infant's body (Stern, 1985). This strategy has the same characteristics of adult behavior described in Sect. 5.5. During intersubjective encounters in which the caregivers try to decrease the infant's alertness, their behavior is characterized by being multimodal and temporally modeled, in this case in terms of rhythmic patterns. Thus, it seems reasonable to assert that there can only occur changes in the infant's general level of arousal or alertness as a consequence of multimodal and dynamically modeled stimulation by means of rhythm if the infant possesses the capacity to detect multimodal rhythmic patterns.

Koester et al. (1989) conducted a study in which they analyzed the first three minutes of a social play session between 3-month-old infants and their mothers. They found that rhythmic movements cover the largest proportion of adult behavior during the interaction. They also observed that the infant's visual attention (looking at mom or in another direction) changed as a function of the mother's movement patterns. The authors concluded that the mothers observed in the study monitored alertness and responded accordingly with variations in the kind of behavior used to redirect or maintain their infant's attention.

Moreno-Núñez et al. (2015) observed the interaction of three mothers with their respective infants, focusing on the communicative mediators of adults (demonstrations and ostensive gestures). One interesting finding from this study is that at 2 months of age of the infants, a significant part of the mothers' ostensive gestures was performed by means of rhythmic movements. Below, as an example, we transcribe the observational account (Observation 1) of Moreno-Núñez et al. (2015) involving a mother and her baby (identified as Child3).

> Child3 0; 1, 16[3] [duration 54 s]. The adult makes a distant demonstration of the sonorous ring between herself and the child. Child3 is reclined on the sofa while the adult holds her by the arm. Child3 constantly looks at the adult's face, smiling at times. The adult speaks with a melodic intonation: "Look how beautiful this iiis…" placing a non-sonorous ring between herself and Child3, moving it rhythmically. "Loook… Oh, how beautiful this iiis…!" intonating the words as descending *glissandos* and stroking her index finger down Child3's cheek in the same direction. Child3 looks at the object and then smiles at and babbles the adult back. The adult continues showing her the object, accompanying this action with descending *glissandos* ("Aaaaaaaah…"), according to the girl's babble. Child3 becomes a bit restless, so the adult releases the non-sonorous ring and *picks up the sonorous one, shaking* it in Child3's face and *recovering her attention*. After a *few repetitions*, she accompanies the *glissandos* with *rhythmic uses of the ring, shaking it to the rhythm of the vocalizations emitted by the child.* (Moreno-Núñez et al., 2015, pp. 744–745) (emphasis in original)

[3] According to the Piagetian convention: year; month (day).

This observation highlights how the rhythmic pattern with which the mother presents the object (the sound ring) attracts the infant's visual attention, promoting its alternation between the mother's face and the object in question.

Ospina (2010)[4] analyzed the interaction between a father and his 5-month-old daughter from the Guapi community.[5] In this case, the observation allows to identify how the detection of the rhythmic pattern that models her father's behavior allows the baby to discover new possibilities in her movements (Ospina & Español, 2014). During the interaction, the father makes a rhythmic offer and the baby is sensitively disposed toward it, in such a way that her body becomes posturally permeable to allow the stimulation to fully impact her. Below we transcribe the observational account analyzed in Ospina (2010):

> The father, holding the baby by her armpits, begins to shake her upward several times so that Faniela rises a little and falls. These jerks are serial, continuous, rapid, energetic, have an underlying pulse, and each is accompanied by a vocalization from the father (cha-cha-cha) and a small tap on the baby's back with each fall. At the beginning of the interaction the baby adjusts her posture, she is observed to be upright but firm and flexible to each of the shocks, allowing the impact of the rebound on her whole body. From the beginning, she is in tune with the father's stimulation proposal, and even though she does not look at him at first, from the very first jolt we can see a joyful expression on her face. […] During this first period, the baby does not experience the agency of these movements, but she does experience her father's attention on her, the encounter between what the father offers and her receptiveness to it, and the strength she exerts to hold the object in her hands. It is worth stressing that this object, a rattle, sounds with some of the shakes and bounces on the father's hand. We observed some attempts on the part of the baby to move the rattle, but it is in the nineteenth shake that she takes a strong impulse with her left arm and manages to lift the rattle, thus beginning a series of arm shakes with it. The first four shakes of the baby overlap with the father's shakes, but when the father stops, the baby continues with her own force keeping the pulse with which the father started, so that a continuity is maintained in the qualities of speed (suddenness), pulse and form (raising-sinking), between the father's and the baby's shakes. (Ospina 2010, pp. 144–145)

According to Ospina and Español (2014), the emerging rhythmic organization between adult and baby allows the baby to discover new possibilities in her movements. It is possible to identify in the baby two ways of appropriation of the rhythmic elements and movement qualities that model the stimulation offered by the father. First, it allows her to take up exactly the pulse and the qualities of the father's stimulation in her arm movements during the joint movement. Second, it allows her to reproduce exactly the pulse and energy qualities of the father's stimulation with

[4] Ospina, V. (2010). El bebé, su condición sensible y la constitución de sí mismo: un avistamiento desde su movimiento en el encuentro con el adulto [The baby, its sensitive condition and the constitution of itself: a glimpse from its movement in the encounter with the adult]. Unpublished master's thesis. Facultad Latinoamericana de Ciencias Sociales – Universidad Autónoma de Madrid.

[5] The municipality of Guapi is located in the southwest of the department of Cauca, bordering the Colombian Pacific slope on the banks of the Guapi River, at an altitude of 5 meters above sea level and an average temperature of 29 °C. The population of Guapi is estimated at 30,000 habitants (guapireños). The main economic activities are small-scale fishing, mining, and agriculture.

her own arm in a post-stimulation manner and thanks to the pauses offered by the father.

As in the case of physiological changes, we can assert that the behavioral changes (in visual orientation, for example) that are found in intersubjective episodes (with and without objects), in which caregivers offer multimodal information dynamically modeled through rhythm, can only occur if the infant possesses the ability to detect them. In this section, we have established some reasonable relationships between the development of multimodal rhythmic pattern detection and some changes in the infant's social behavior.

In the examples presented above, it can be observed how caregivers produce changes in the infant's behavior by regulating his/her emotional state or mood, orienting his/her visual attention toward the interaction partner or toward different objects that adults usually introduce during intersubjective encounters, and making it possible for the infant to "agentively" add new qualities to the movements he/she performs. We believe that these changes in the infant's social behavior (social in that they occur as a function of intersubjective experience) depend on his/her ability to detect multimodal rhythmic patterns.

In this section, we have dealt with the development of perceptual skills and changes in the infant's social behavior that occur at a very early period, the last months of gestation and the first five months after birth. In the following sections, we address the perceptual skills and changes in the infant's social behavior at much later stages, from nine months onward. We are fully aware that the time gap is quite large, especially if our focus is on the first year of life. Surely, during this period, several changes in the infant's social behavior must occur, but we have not found in the literature any observational accounts that would allow us to carry out an analysis as the one we are proposing. We think it is imperative to undertake a systematic study of the changes that occur during this period in the infant's social behavior (from five to nine months) concerning the detection of rhythmic patterns in order to fully understand the relation between the development of multisensory perception and intersubjective development.

5.7.2 Discrimination of Rhythmic Patterns and the Intersubjective Encounter

Recall that discrimination refers to the ability to differentiate between two rhythmic patterns. As previously reviewed (Sect. 5.6), from the fourth month onward infants begin to discriminate multimodal rhythmic patterns and from the eighth month onward unimodal rhythmic patterns. The ability to discriminate different rhythmic patterns may be closely linked to the development of the ability to segment the perceptual flow coming from the behavior of the other during an interaction. According to Español (2004), the repetition of a rhythmic sequence allows the infant to decompose (*discretize*) the other's action and anticipate what is to come. Now, why is the

ability to discriminate between different rhythmic patterns fundamental to acquiring this skill? Quite simply, we believe it is fundamental because the repetition-variation form is constitutive of adult performance (Español & Shifres, 2015, see Chap. 2 in this volume). Recall that under this format, the adult provides the infant with stimulation that is repetitive and stable enough for him to perceive clearly, but also variable enough to sustain his attention. If we think of a prototypical performance in which the varied repetition falls on the rhythmic pattern, it is clear that the ability to discriminate between different rhythmic patterns plays a decisive role in discretizing the adult's behavior.

To show how the development of multisensory perception is related to the change in the infant's interactive behavior, we will transcribe two observations (Observation 1 y 2) made by Español (2004) where we see how a 9-month-old child segments and anticipates the behavior of the adult with whom he is interacting when she performs two different performances, one multimodal and the other unimodal.

> When H. is 0; 9 (02), I begin one of those typical interactions consisting of singing while moving my hands. H. is attentive all the time looking at my hands or making eye contact. When I finish H. keeps looking at me and then I repeat the game. And so on for several times. After a while I approach him, little by little, producing rhythmic sounds until I bump my head against his. H. looks at me with interest and the scene goes on with continuing eye contact. Little by little he begins to bring his head closer to mine when the moment of the head-to-head bump arrives. Just when the time comes to bring my head closer I delay the movement and it is H. who initiates it. The sequence is repeated with H. initiating the approach movement a few more times. (Español, 2004, p. 59)

> At 0; 9 (25), H. looks at me while I make rhythms and hand movements. He just looks at me, then goes back to his items. I stretch out my foot and touch H, he smiles, I grab him with my legs and bring him toward me and forward with my legs several times, while singing. H. takes my feet, looks at them; he looks very interested, I move them rhythmically. H. does not let go of his feet, he accompanies the movement with his hands. When I stop moving my feet, he begins to move his. We don't look at each other, our backs are turned. I cross my legs. H. stares at my foot, takes it and moves it sideways. I restart the game. Several times like this. The whole scene lasts about four minutes. (Español, 2004, p. 59)

Observation 1 shows how the infant anticipates the behavior of the adult while she displays a multimodal performance, an interaction in which there is persistent eye contact. Observation 2 also shows how the infant anticipates the adult's behavior, but in this case it is a performance in which, at a particular moment, the only source of stimulation (hence unimodal) the infant receives is kinesthetic information when his body is moved by the adult's feet and without eye contact. These are two more typical examples of plausible relations between the infant's mode of engagement during the interaction and his/her multisensory perceptual abilities, specifically based in this case on the ability to discriminate rhythmic patterns. In the first example, the anticipation of the adult's behavior occurs as the infant participates in an intersubjective encounter in which the repetition with variations falls on the rhythmic modeling of a multimodal performance. In the second example, the anticipation of the adult's behavior occurs while the infant participates in an intersubjective encounter in which the repetition with variations falls on the rhythmic modeling of a unimodal performance.

5.8 Recognition of Multimodal Rhythmic Patterns and the Intersubjective Encounter

Recall that recognition refers to the ability to establish relationships between two amodal features of an event when they occur simultaneously, as, for example, in the rhythmic modeling of movement and sound present in the mother's behavior during intersubjective encounters.

At the beginning of Sect. 5.3, we mentioned that most studies on early intersubjective development focus on the ninth-month revolution or the transition from primary to secondary intersubjectivity. We claimed that this is not the only change that needs to be explained. Consequently, we have given an account of how other skills involved in intersubjective contact can be reasonably linked to the development of multisensory perception. However, it also seems important to us to account, in part, for the transition from one mode of intersubjectivity to another. One of the skills that emerges during the period of secondary intersubjectivity is the visual alternation between infants and adults as they share their experience about an object or an event. Again, in order to establish a reasonable relationship between the development of this skill and the development of multisensory perception, we will refer to a fragment of an observational account from a microanalysis of the interaction between a mother and her 10-month-old daughter in a social play situation. The observation is taken from Fogel and DeKoeyer-Laros (2007):

> Susan glances at the table and then looks at Sheryl's face. Raising her eyebrows, she starts to pound. Pounding the table top with flat hands, she follows Sheryl's movements with her gaze. It appears as though she wants to initiate the pounding game. But Sheryl says, "I wanna show off and show how you can do pat-a-cake[6]." Susan stops and reaches for Sheryl's face. Sheryl starts to sing and clap 'pat-a-cake'. Continuing to stare intently at her singing and clapping mother, Susan starts to pound the table again. *Continuing to pound, she shifts her gaze to her hands and then back to Sheryl*, who says, "No, show them how you clap, okay?" (Fogel & DeKoeyer-Laros, 2007, p. 80) (emphasis in original)

The fragment of the observation shows how the child initiates the interaction sequence and performs the alternation of gaze between her tapping and the mother's face and tapping. The interaction described is not a spontaneous game between the mother and the little daughter. It responds to the play that, historically, they were developing while participating in the research that described their behaviors. This is an example of the so-called historical performances (see Chap. 3). Fogel and DeKoeyer-Laros (2007) argue that the mother scaffolded the multisensory tapping co-actions in order to stabilize the rhythmic hand movements of both participants, allowing the child to alternate her gaze between her tapping and her mother's behavior and face. It is only possible for the child to interact with her mother (while performing a multisensory co-action) if she has the ability to recognize that the rhythm of her tapping on the tabletop is the same as that of her mother's tapping. Indeed,

[6] Traditional English song that adults often sing to babies during early social play, usually accompanied by hand clapping.

this is because at 10 months infants recognize intersensory relationships based on rhythm (see Sect. 6.3). Again, we have been able to identify in an interactive situation the link between the development of multisensory perception and early intersubjective development.

5.9 Final Remarks

In this chapter, we have attempted to relate the available empirical data on the development of multisensory perception (detection, discrimination, and recognition) with the available observational data on the infant's participation in social interactions where the adult's behavior is rhythmically and multimodally modeled. The relationship we have sought to establish goes beyond the identification of the synchronous occurrence of both phenomena. The analyses carried out have allowed us to identify those aspects of the rhythmic structure of the adult's behavior that only when they are detected, discriminated, or recognized allow a particular behavior of the infant to take place that would not occur if the infant had not developed any one of these multisensory perception skills.

Our analysis has focused solely on rhythm, since as we have already mentioned, rhythm is an essential quality of experience, is present in intersubjective exchanges from the very beginning of life, and has been the object of special attention in research on early development. However, more studies are needed to provide relevant information to extend the link between intersubjective development and the development of multisensory perception in terms of the other two types of temporal amodal information (duration and rate) that dynamically shape adult behavior.

The finding of plausible relationships between intersubjective development and the development of multisensory perception allows us to interpret some relevant changes in the infant's interactive behavior in clear harmony with the view that the infant's social development emerges progressively beyond the transition from primary to secondary intersubjectivity (Rivière, 1986/2003). In this regard, we believe that the analysis carried out in this chapter allows us to address a somewhat elusive aspect in the philosophical and psychological works under the umbrella of the SPP. It is important to emphasize that in most of the works listed under this approach there is a remarkable absence of theoretical models that systematically link early intersubjective development with the development of perception, especially as regards the important role of perception in the establishment of intersubjective contact. This absence is probably due to the scarcity of data and analyses such as those we have undertaken here.

In this sense, an essential contribution of the data presented in Sect. 5.7 is that they allow formulating new research questions that will pave the way for observational studies aimed at understanding early intersubjective development, both in dyadic settings of early social play and in object-mediated interactive situations, in close connection with the development of multisensory perception.

References

Abramova, E., & Slors, M. (2015). Social cognition in simple action coordination: A case for direct perception. *Consciousness and Cognition: An International Journal, 36*, 519–531. https://doi.org/10.1016/j.concog.2015.04.013

Bahrick, L. E., & Lickliter, R. (2000). Intersensory redundancy guides attentional selectivity and perceptual learning in infancy. *Developmental Psychology, 36*, 190–201. https://doi.org/10.1037/0012-1649.36.2.190

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective and Behavioral Neuroscience, 4*, 137–147. https://doi.org/10.3758/CABN.4.2.137

Baillargeon, R., & Carey, S. (2012). Core cognition and beyond: The acquisition of physical and numerical knowledge. In S. M. Pauen (Ed.), *Early childhood development and later outcome* (pp. 33–65). Cambridge University Press.

Baron Cohen, S. (2005). The empathizing system: A revision of the 1994 model of the mindreading system. In B. Ellis & D. Bjorklund (Eds.), *Origins of the social mind evolutionary psychology and child development* (pp. 468–492). Guilford Press.

Baron-Cohen, S., Tager-Flusberg, H., & Lombardo, M. (2013). *Understanding others minds. Perspectives from developmental social neuroscience*. Oxford University Press.

Barone, P., & Gomila, A. (2020). Infants' performance in the indirect false belief tasks: A second-person interpretation. *Wiley Interdisciplinary Review: Cognitive Science, 12*, e1551. https://doi.org/10.1002/wcs.1551

Barone, P., Corradi, G., & Gomila, A. (2019). Infants' performance in spontaneous-response false belief tasks: A review and meta-analysis. *Infant Behavior & Development, 57*, 101350. https://doi.org/10.1016/j.infbeh.2019.101350

Bates, E. (1976). *Language and context. The acquisition of performatives*. Academic Pres.

Bates, E., Camaioni, L., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21*(3), 205–226.

Bateson, M. C. (1979). The epigenesis of conversational interpersonal: A personal account of research development. In M. Bullowa (Ed.), *Before speech. The beginning OS interpersonal communication* (pp. 63–78). Cambridge University Press.

Bohl, V., & van den Bos, W. (2012). Toward an integrative account of social cognition: Marrying theory of mind and interactionism to study the interplay of type 1 and type 2 processes. *Frontiers in Human Neuroscience, 6*, 274. https://doi.org/10.3389/fnhum.2012.00274

Bordoni, M., Español, S., & De Grande, P. (2016). La incidencia del entonamiento afectivo y la imitación en el involucramiento visual-social temprano [The incidence of affect attunement and imitation in early social-visual engagement]. *Avances en Psicología Latinoamericana, 34*(3), 487–503. https://doi.org/10.12804/apl34.3.2016.04

Bower, T. G. R. (1979). *Human development* [Psicología del desarrollo (A. Rosa, S. Fae-Peters, P. del Río y J. del Pozo, trads.; A. Rosa rev. tec.). Madrid: Siglo XXI, 1983]. Freeman and Company.

Bowlby, L. (1982). *Attachment and loss. I Attachment* [El apego y la pérdida. 1 – El apego (M. Valcarce Avello, trad). Buenos Aires: Paidós, 1998] (2nd ed). The Hogath Press.

Bråten, S. (1998). Introducction. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 1–12). Cambridge University Press.

Bruner, J. (1983). *Child´s talk. Learning to use language* [El habla del niño (R. Premat, trad.; I. Vila, sup.) Barcelona: Paidós, 1986]. W. W. Norton.

Carey, S. (2011). The origin of concepts: A précis. *Behavioral and Brain Sciences, 34*, 113–167. https://doi.org/10.1017/S0140525X10000919

Carretero, S., & Español, S. A. (2016). Multimodal study of adult-infant interaction: A review of its origins and its current status. *Paidéia (Ribeirão Preto), 26*(65), 377–385. https://doi.org/10.1590/1982-43272665201613

Carruthers, P., & Smith, P. K. (Eds.). (1996). *Theories of theories of mind*. Cambridge University Press. https://doi.org/10.1017/CBO9780511597985

Collis, G. M., & Schaffer, H. R. (1975). Synchronization of visual attention in mother-infant pairs. *Child Psychology & Psychiatry & Allied Disciplines, 16*(4), 315–320. https://doi.org/10.1111/j.1469-7610.1975.tb00365.x

Davis, J., Redshaw, J., Suddendorf, T., et al. (2021). Does neonatal imitation exist? Insights from a meta-analysis of 336 effect sizes. *Perspectives on Psychological Science, 16*(6), 1373–1397. https://doi.org/10.1177/1745691620959834

de la Paz Jacquier, M., Martínez, G., Pereira Ghiena, A., & Silva, V. (2013). La organización rítmica de la música [The rhythmic organization of music]. In F. Shifres & M. I. Burcet (Eds.), *Escuchar y pensar la música: bases teóricas y metodológicas* [Listening and thinking about music: Theoretical and methodological basis] (pp. 172−204). EDULP.

Dissanayake, E. (2014). Homo musicus: ¿Estamos biológicamente predispuestos para ser musicales [Homo musicus: Are we biologically predisposed to be musical?]. In: S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality] (pp. 195−215). Paidós.

Eckerdal, P., & Merker, B. (2009). "Music" and the "action song" in infant development: An interpretation. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality: Exploring the basis of human companionship* (pp. 241–258). Oxford University Press.

Ekman, P. (1982). *Emotions in the human face*. Cambridge University Press.

Elman, J. L., Bates, E. A., Johnson, M. H., Karmiloff-Smith, A., Parisi, D., & Plunkett, K. (1996). *Rethinking innateness: A connectionist perspective on development*. The MIT Press.

Emery, N. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neuroscience and Biobehavioral Reviews, 24*(6), 581–604. https://doi.org/10.1016/s0149-7634(00)00025-7

Español, S. (2003). De la emoción al espíritu metafórico. Semiosis e intersubjetividad en el desarrollo humano [From emotion to metaphorical spirit: Semiosis and intersubjectivity in human development]. *Studies in Psychology, 24*(3), 277–311. https://doi.org/10.1174/021093903770411193

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana* [How to do things without words. Gesture and fiction in early infancy]. Antonio Machado Libros.

Español, E. (2005). Ontogénesis de la experiencia estética. La actitud contemplativa y las artes temporales en la infancia [Ontogenesis of aesthetic experience: Contemplative attitude and temporal arts in infancy]. *Studies in Psychology, 26*(2), 139–171. https://doi.org/10.1174/0210939054024885

Español, S. (2014). La forma repetición-variación: una estrategia para la reciprocidad [The repetition-variation form: a strategy for reciprocity]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality] (pp. 157−192). Paidós.

Español, S., & Shifres, F. (2015). The artistic infant directed performance: A microanalysis of the adult's movements and sounds. *Integrative Psychological and Behavioral Science, 49*(3), 371–397. https://doi.org/10.1007/s12124-015-9308-4

Español, S., Bordoni, M., Carretero Pérez, S., Martínez, M., & Camarasa, R. (2018). Imitación y entonamiento afectivo en el juego social temprano [Imitation and affect attunement in early social play]. *Interdisciplinaria: Revista de Psicología y Afines, 35*(2), 291–305.

Fabb, N., & Halle, M. (2011). Grouping in the stressing of words, in metrical verse, and in music. In P. Rebuschat, M. Rohmeier, J. A. Hawkins, & I. Cross (Eds.), *Language and music as cognitive systems* (pp. 4–21). Oxford University Press.

Fantz, R. L. (1958). Pattern vision in young infants. *The Psychological Record, 8*, 43–47.

Feldman, R., Magori-Cohen, R., Galili, G., Singer, M., & Louzoun, Y. (2011). Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development, 34*(4), 569–577. https://doi.org/10.1016/j.infbeh.2011.06.008

Field, T. (1978). The three Rs of infant-adult interactions: Rhythms, repertoires, and responsibility. *Journal of Pediatric Psychology, 3*(3), 131–136. https://doi.org/10.1093/jpepsy/3.3.131

Fogel, A., & DeKoeyer-Laros, I. (2007). The developmental transition to secondary intersubjectivity in the second half year: A microgenetic case study. *Journal of Developmental Processes, 2*, 63–90.

Freud, S. (1940). Abriss der Psychoanalyse [Esquema del psicoanálisis (J. L. Etcheverry, trad.). Buenos Aires: Amorrortu, 2010]. *Internationale Zeitschrift für Psychoanalyse–Imago*, *25*(1), 7–67.

Gallagher, S. (2008). Direct perception in the intersubjective context. *Consciousness and Cognition, 17*(2), 535–543. https://doi.org/10.1016/j.concog.2008.03.003

Gergely, G., & Csibra, G. (2003). Teleological reasoning in infancy: The naïve theory of rational action. *Trends in Cognitive Sciences, 7*(7), 287–292. https://doi.org/10.1016/S1364-6613(03)00128-1

Gómez, J. C. (1998). Do concepts of intersubjeticity apply to non-human primates? In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 245–259). Cambridge University Press.

Gómez, J. C. (2008). The evolution of pretense: From intencional availability to intencional non-existence. *Mind and Language, 23*(5), 586–606. https://doi.org/10.1111/j.1468-0017.2008.00353.x

Gómez, J. C. (2009). Embodying meaning: Insights from primates, autism, and Brentano. *Neural Networks, 22*, 190–196. https://doi.org/10.1016/j.neunet.2009.01.010

Gómez, J. C. (2010). L´evoluzione del mentalismo intenzionale: della disponibilitá intenzionale alle intenzioni sul nulla [The evolution of intentional mentalism: From intentional availability to intentions about nothing]. *Metis, 17*, 7–27.

Gopnik, A., & Wellman, H. M. (1994). The theory. In L. Hirschfeld & S. Gelman (Eds.), *Domain specificity in cognition and culture* (pp. 257–293). Cambridge University Press.

Gray, H. (1978). Learning to take an object from the mother. In A. Lock (Ed.), *Action, gesture and symbol* (pp. 158–181). Academic Press.

Hobson, R. P. (1993). *Autism and the development of mind* [El autismo y el desarrollo de la mente (A. Rivière, trad.). Madrid: Alianza Editorial, 1995]. LEA Ltd.

Hubley, P., & Trevarthen, C. (1979). Sharing a task in infancy. In I. Uzgiris (Ed.), *Social interaction during infancy: New directions for child development* (Vol. 4, pp. 57–80). Jossey Bass.

Iacoboni, M. (2008). *Mirroring people. The new science of how we connect with others* [Las neuronas espejo: Empatía, neuropolítica, autismo, imitación, o de cómo entendemos a los otros (I. Rodríguez Villegas, trad.). Madrid: Katz Editores, 2009]. Farrar, Straus and Giroux.

Jackendoff, R. S. (2002). *Foundations of language: Brain, meaning, grammar, evolution*. Oxford University Press.

James, W. (1890). *The principles of psychology* [Principios de psicología (D. Barnes, trad.). Buenos Aires: Editorial GLEM, 1945]. Henry Holt and Company.

Karmiloff-Smith, A. (2015). An alternative to domain-general or domain-specific frameworks for theorizing about human evolution and ontogenesis. *AIMS Neuroscience, 2*(2), 91–104. https://doi.org/10.3934/Neuroscience.2015.2.91

Kaye, K. (1982). *The mental and social life of babies: How parents create persons* [La vida mental y social del bebé. De cómo los padres crean personas (D. Rosenbaum, trad., C. Coll, sup.). Barcelona: Paidós, 1986]. University Chicago Press.

Koester, L. S., Papoušek, H., & Papoušek, M. (1989). Patterns of rhythmic stimulation by mothers with three-month-olds: A cross-modal comparison. *International Journal of Behavioral Development, 12*(2), 143–154. https://doi.org/10.1177/016502548901200201

Korner, A. F., & Thoman, E. B. (1970). Visual alertness in neonates as evoked by maternal care. *Journal of Experimental Child Psychology, 10*(1), 67–78. https://doi.org/10.1016/0022-0965(70)90045-7

Krueger, J. (2018). Direct social perception. In A. Newen, L. De Bruin, & S. Gallagher (Eds.), *The Oxford handbook of 4E cognition* (pp. 301–320). Oxford University Pres.

Lany, J., & Saffran, J. R. (2013). Statistical learning mechanisms in infancy. Neural Circuit Development and Function in the Brain. *Comprehensive Developmental 231 Neuroscience, 3*. https://doi.org/10.1016/B978-0-12-397267-5.00034-0

Lecanuet, J. P., & Jacquet, A. Y. (2002). Fetal responsiveness to maternal passive swinging in low heart rate variability state: Effects of stimulation direction and duration. *Developmental Psychobiology: The Journal of the International Society for Developmental Psychobiology, 40*(1), 57–67. https://doi.org/10.1002/dev.10013

Leslie, A. M. (1987). Pretense and representation: The origins of "theory of mind.". *Psychological Review, 94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*(2), 281–308. https://doi.org/10.1037/0033-2909.126.2.281

Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology, 39*(5), 795. https://doi.org/10.1037/0012-1649.39.5.795

Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4-to 10-month-old infants. *Developmental Psychobiology, 48*(4), 288–300. https://doi.org/10.1002/dev.20140

Malloch, S., & Trevarthen, C. (2009). *Communicative musicality: Exploring the basis of human companionship*. Oxford University Press.

Martínez, I. (2014). La base corporeizada del significado musical [The embodied basis of musical meaning]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality] (pp. 71−110). Paidós.

Martínez, M., Español, S., & Igoa, J. M. (2018). Reconocimiento de relaciones intersensoriales basadas en el ritmo a los 4, 7 y 10 meses [Recognition of rhythm–based intersensory relationships at 4, 7, and 10 months]. *Epistemus, Revista de Estudios en Música, Cognición y Cultura, 6*(2), 10–32. https://doi.org/10.21932/epistemus.6.6233.2

Melinder, A., Forbes, D., Tronick, E., Fikke, L., & Gredebäck, G. (2010). The development of the still-face effect: Mothers do matter. *Infant Behavior and Development, 33*(4), 472–481. https://doi.org/10.1016/j.infbeh.2010.05.003

Meltzoff, A. N. (2011). Social cognition and the origins of imitation, empathy, and theory of mind. In U. Goswami (Ed.), *The Wiley-Blackwell handbook of childhood cognitive development* (2nd ed., pp. 49–75). Wiley-Blackwell.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science, 198*, 75–78. https://doi.org/10.1126/science.897687

Meltzoff, A. N., & Moore, M. K. (1997). Explaining facial imitation: A theoretical model. *Early Development and Parenting, 6*(3-4), 179–192.

Meltzoff, A. N., & Moore, M. K. (1998). Infant intersubjectivity: Broadening the dialogue to include imitation, identity and intention. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 47–62). Cambridge University Press.

Meltzoff, A. N., & Williamson, R. A. (2017). Imitation and modeling. In M. M. Haith & J. B. Benson (Eds.), *Encyclopedia of infant and early childhood development: Reference module in neuroscience and biobehavioral psychology* (pp. 1–10). Elsevier.

Miller, N. E., & Dollard, J. (1941). *Social learning and imitation*. Yale University Press.

Moreno-Núñez, A., Rodríguez, C., & Del Olmo, M. J. (2015). The rhythmic, sonorous and melodic components of adult-child-object interactions between 2 and 6 months old.

Integrative Psychological and Behavioral Science, 49(4), 737–756. https://doi.org/10.1007/s12124-015-9298-2

Mundy, P. (2016). *Autism and joint attention: Development, neuroscience and clinical fundamentals*. Guilford Publications, Inc.

Murillo, E., Montero, I., & Casla, M. (2021). On the multimodal path to language: The relationship between rhythmic movements and deictic gestures at the end of the first year. *Frontiers in Psychology, 12*, 616812. https://doi.org/10.3389/fpsyg.2021.616812

Nespor, M., & Vogel, I. (1986). *Prosodic phonology* [La prosodia (A. Ardid Gumiel, trad.). Madrid: Visor, 1994]. Foris Publications.

Nooteboom, S. G. (1997). The prosody of speech: Melody and rhythm. In W. J. Hardcastle & J. Laver (Eds.), *The handbook of phonetic sciences* (pp. 640–673). Blackwell.

Onishi, K. H., & Baillargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science, 308*(5719), 255–258. https://doi.org/10.1126/science.1107621

Osler, L. C. (2016). Is direct social perception a banal thesis? A phenomenological response to Spaulding's critique of Direct Social Perception. *Tidsskrift for Medier, Erkendelse og Formidling, 4*(2), 775–793.

Ospina, V., & Español, S. (2014). El movimiento y el sí mismo [Movement and the self]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality] (pp. 111−155). Paidós.

Papoušek, H. (1996). Musicality in infancy research: Biological and cultural origins of early musicality. In I. Deliege & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 37–55). Oxford University Press.

Papoušek, M., & Papoušek, H. (1981). Musical elements in the infant´s vocalizations: Their significance for communication, cognition and creativity. In L. P. Lipsitt (Ed.), *Advances in infancy research* (Vol. 1, pp. 163–224). Ablex Norwood.

Pauen, M. (2012). The second-person perspective. *Inquiry, 55*(1), 33–49. https://doi.org/10.1080/0020174x.2012.643623

Pérez, D., & Gomila, A. (2018). La atribución mental y la segunda persona [Mental attribution and the second person]. In T. Balmaceda & K. Pedace (Comps.), *Temas de filosofía de la mente. Atribución psicológica* [Topics in philosophy of mind. Psychological attribution] (pp. 69−98). SADAF.

Pérez, D., & Gomila, A. (2022). *Social cognition and the second person in human interaction*. Routledge.

Perner, J. (1991). *Understanding the representational mind* [Comprender la mente representacional (M. A. Galmarini, trad.; E. Martí, rev. tec.). Barcelona: Paidós, 1984]. The MIT Press.

Pickens, J. N., & Bahrick, L. E. (1995). Infants' discrimination of events on the basis of rhythm and tempo. *British Journal of Developmental Psychology, l3*, 223–236. https://doi.org/10.1111/j.2044-835X.1995.tb00676.x

Pickens, J., & Bahrick, L. E. (1997). Do infants perceive invariant tempo and rhythm in auditory-visual events? *Infant Behavior and Development, 20*(3), 349–357. https://doi.org/10.1016/S0163-6383(97)90006-0

Premack, D., & Woodruff, G. (1978). Does the chimpanzee have a theory of mind? *Behavioral and Brain Sciences, 1*(4), 515–526. https://doi.org/10.1017/S0140525X00076512

Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in Psychology, 5*, 1048. https://doi.org/10.3389/fpsyg.2014.01048

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Rivière, A. (1986/2003). Interacción precoz. Una perspectiva vygotskiana a partir de los esquemas de Piaget [Early interaction. A Vygotskian perspective from Piaget's schemas]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comp.) *Ángel Rivière. Obras Escogidas* [Ángel Rivière. Selected Works], *Vol II* (pp. 109–142). Panamericana.

Rochat, P. (2001). *The infants world* [El mundo del bebé (R. Filella, trad.). Madrid: Morata, 2004]. Harvard University Press.

Rowlands, M. J. (2006). *Body language*. The MIT Press.

Rowlands, M. J. (2010). *The new science of the mind: From extended mind to embodied phenomenology*. The MIT Press.

Saby, J. N., Marshall, P. J., & Meltzoff, A. N. (2012). Neural correlates of being imitated: An EEG study in preverbal infants. *Social Neuroscience, 7*, 650–661. https://doi.org/10.1080/17470919.2012.691429

Sarriá, E., & Rivière, Á. (1991). Desarrollo cognitivo y comunicación intencional preverbal: un estudio longitudinal multivariado [Cognitive development and preverbal intentional communication: A multivariate longitudinal study]. *Studies in Psychology, 46*, 35–52. https://doi.org/10.1080/02109395.1991.10821160

Schaffer, R. H. (1971). The growth of sociability [El desarrollo de la sociabilidad (E. La Fuente, trad.). Madrid: Aprendizaje Visor, 1983]. Penguin Books Ltd.

Schaffer, R. H. (1977). Early interactive development. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: Proceedings of the Loch Lomond symposium* (pp. 3–16). Academic Press.

Scotto, C. (2017). Lo que el aprendizaje del lenguaje revela sobre el lenguaje (y sobre la cognición social) [What language learning reveals about language (and about social cognition)]. In D. Pérez and D. Lawler (Eds.), *La segunda persona y las emociones* [The second person and emotions] (pp. 87−140). SADAF.

Sears, R. R. (1951). A theoretical framework for personality and social behavior. *American Psychologist, 6*(9), 476–482. https://doi.org/10.1037/h0063671

Spaulding, S. (2015). On direct social perception. *Consciousness and Cognition, 36*, 472–482. https://doi.org/10.1016/j.concog.2015.01.003

Stern, D. N. (1971). A micro-analysis of mother-infant interaction: Behavior regulating social contact between a mother and her 3 1/2-month-old twins. *Journal of the American Academy of Child Psychiatry, 10*(3), 501–517. https://doi.org/10.1016/S0002-7138(09)61752-0

Stern, D. N. (1977). *The first relationship: Infant and mother* [La primera relación madre-hijo (A. Guerra Millares, trad.). Madrid: Morata, 1983]. Harvard University Press.

Stern, D. N. (1985). The interpersonal world of the infant. A view from Psychoanalysis and development psychology [El mundo interpersonal del infante. Una perspectiva desde el psicoanálisis y la psicología evolutiva (J. Piatigorsky, trad.). Buenos Aires: Paidós, 1991]. Basic Books, Inc.

Stern, D. N. (2010). *Forms of vitality: Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford University Press.

Stern, D., Hofer, L., Haft, W., & Dore, J. (1985). Affect attunement: The sharing of feeling states between mother and infant by means of intermodal fluency. In T. Field & N. Fox (Eds.), *Social perception in infants* (pp. 249–268). Ablex.

Tomasello, M. (1999). *The cultura origins of human cognition* [Los orígenes culturales de la cognición humana (A. Negrotto, trad.). Buenos Aires: Amorrortu, 2007]. Harvard University Press.

Tomasello, M. (2008). *Origins of human communication* [Los orígenes de la comunicación humana (E. Marengo, trad.). Barcelona: Katz Editores, 2013]. MIT Press.

Trevarthen, C. (1974). Conversations with a two-month-old. *New Scientist, 62*(2), 230–235.

Trevarthen, C. (1977). Descriptive analysis of infant communicative behaviour. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction: Proceedings of the Loch Lomond symposium* (pp. 227–270). Academic Press.

Trevarthen, C. (1980). The foundations of intersubjectivity: Development of interpersonal and cooperative understanding in infants. In D. Olson (Ed.), *The social foundations of language and thought. Essays in honor of Jerome Bruner* (pp. 316–342). W.W. Norton & Co..

Trevarthen, C. (1982). Los motivos primordiales para entenderse y para cooperar [The primary motives for cooperative understanding]. In Perinat, A (Ed.), *La comunicación preverbal* [Preverbal communication] (pp. 77−109). Ediciones Avesta.

Trevarthen, C. (1998). The concept and foundations of infant intersubjectivity. In S. Bråten (Ed.), *Intersubjective communication and emotion in early ontogeny* (pp. 15–46). Cambridge University Press.
Trevarthen, C. (2001). Intrinsic motives for companionship in understanding: Their origin, development and significance for infant mental health. *Infant Mental Health Journal, 22*(1-2), 95–131.
Trevarthen, C. (2015). Infant semiosis: The psycho-biology of action and shared experience from birth. *Cognitive Development, 36*, 130–141. https://doi.org/10.1016/j.cogdev.2015.09.008
Trevarthen, C. (2017). Play with infants: The impulse for human story-telling. In T. Bruce, P. Hakkarainen, & M. Bredikyte (Eds.), *The Routledge international handbook of play in early childhood* (pp. 198–215). Taylor & Francis/Routledge.
Trevarthen, C., & Hubley, P. (1978). Secondary intersubjectivity: Confidence, confiding and acts of meaning in the first year. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 183–229). Academic Press.
Trevarthen, C., & Reddy, V. (2007). Consciousness in infants. In M. Velmansy & S. Schneider (Eds.), *The Blackwell companion to consciousness* (pp. 41–57). Blackwell Publishing.
van der Lely, H. K. (2005). Domain-specific cognitive systems: Insight from Grammatical-SLI. *Trends Cognitive Science, 9*(2), 53–59. https://doi.org/10.1016/j.tics.2004.12.002
Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind. Cognitive science and human experience* [De cuerpo presente. Las ciencias cognitivas y la experiencia humana (C. Gardini, trad.). Barcelona: Gedisa, 2005). MIT Press.
Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*(3), 437. https://doi.org/10.1037/0033-2909.121.3.437
Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development, 72*(3), 655–684. https://doi.org/10.1111/1467-8624.00304
Wiltshire, T. J., Lobato, E. J., McConnell, D. S., & Fiore, S. M. (2015). Prospects for direct social perception: A multi-theoretical integration to further the science of social cognition. *Frontiers in Human Neuroscience, 8*, 1007. https://doi.org/10.3389/fnhum.2014.01007
Winkler, I., Háden, G. P., Ladinig, O., Sziller, I., & Honing, H. (2009). Newborn infants detect the beat in music. *Proceedings of the National Academy of Sciences of the United States of America, 106*(7), 2468–2471. https://doi.org/10.1073/pnas.0809035106

Mauricio Martínez has a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). He participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning at the Faculty of Latin America Social Sciences (FLACSO). He is Professor of Developmental Psychology at Universidad Abierta Interamericana (UAI) and Faculty member of the MsC. in Cognitive Psychology (FLACSO). He also teaches in the post-graduate virtual course "*The baby's life: psychological development, scaffolding and nurturing environmen*" at FLACSO. His research focuses on early perceptual and intersubjective development, childhood play and early intervention in autism spectrum disorders. He is author of *Intervención psicoeducativa para niños con trastorno del espectro autista* [Psychoeducational intervention for children with autism spectrum disorders] (2015). His publications are available at: https://www.researchgate.net/profile/Mauricio-Martinez-2

José Manuel Igoa has a PhD. in Psychology from the Universidad Complutense de Madrid (Spain). His is full Professor at the Department of Basic Psychology at the Universidad Autónoma de Madrid (UAM, Spain). He has focused his activity on the psycholinguistic study of language, the relationship between linguistic processing and other cognitive processes (e.g., perception), and the involvement of domain-specific versus domain-general skills in mental activity in general and language in particular. He is a founding member of the Spanish Society of Experimental Psychology

(Sepex) (1997) and of the UAM-accredited Research Group Cognitive Processes and Language Disorders, Memory and Mentalization. He is co-author, with Mercedes Belinchón and Ángel Rivière, of the book *Psicología del lenguaje. Investigación y teoría* [Psychology of language. Research and theory] (1992). Some recent publications are "Recognition of rhythm-based intersensory relations at 4, 7 and 10 months," "Time-varying effective connectivity during visual object naming as a function of semantic demands" and "Applying language science to language pedagogy", all co-authored with several researchers.

Chapter 6
The Development of Multisensory Perception of Temporal Amodal Information

Mauricio Martínez

Abstract In this chapter, we present a theoretical model of the ontogenesis of multisensory perception of temporal amodal information (MPTAI). The development of MPTAI involves different perceptual abilities (detection, discrimination, and recognition) linked to temporal amodal information (duration, rate, and rhythm). A key aspect of the model is that it assumes a dialectical relationship between early intersubjective development and the development of the perceptual skills it aims to explain. The model attempts to account for the development of MPTAI by combining an ecological approach to perception with a socio-historical-cultural perspective based on the genesis of the infant's exploratory activity that allows him/her to extract the temporal amodal information present in the adult's behavior.

Keywords Intersensory development · Multisensory perception · Rhythm · Social perception · Duration · Rate

6.1 Introduction

Perception is a vast field of research that lacks a comprehensive theory covering the complex nature of this phenomenon. The field is highly fragmented with contributions from biology, physiology, physics, philosophy, and, of course, psychology. Vision is the dominant perceptual system in the field of research on perception (Travieso & Blanco, 1999), and traditionally, each perceptual system has been studied separately. In the research field of infant perceptual development, the picture is quite similar. Rochat argues that "the majority of research on early perception has

M. Martínez (✉)
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina
e-mail: mmartinez@flacso.org.ar

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_6

been performed in the area of vision" (2001, p. 86). Particularly, the topics that have been most widely studied are eye movements, scanning and visual fixation; visual acuity and contrast sensitivity; movement and color perception; depth perception; shape and size constancy; form perception; motion perception; and gestalt organization; among others (Slater et al., 2010). In addition, there are studies directly linked to the perception of objects or people. The latter focus, for example, on face and emotion perception and discrimination (Bremner et al., 2017) and speech perception (Saffran et al., 2006). The study of perception is characterized by adopting a predominantly unisensory approach (i.e., focusing on each of the perceptual systems separately), and providing explanations that deal with very specific aspects (Slater et al., 2010). It seems that infant perception studies are only aimed at identifying at what point infants can perceive the way adults do.

For several years, the study of multisensory perception has offered empirical data and theories about partial aspects of its development. However, it is only recently that this area of concern is attracting the attention of developmental psychologists (Bremner et al., 2012). Thus, the importance of intersensory perception development for social, emotional, cognitive, and linguistic development is beginning to gain understanding and generate research (Bahrick & Lickliter, 2012). A close examination of the studies on multisensory perception offers some regularities. On one hand, we can identify a great variety of studies regarding the ability to perceive multimodal information in a unified way; on the other hand, there are studies that investigate different types of intersensory relationships coming from different sources of information: objects, people and oneself (Martínez, 2016). Studies on the development of multisensory perception seem to fit the general statement that, according to Valsiner (1998), characterizes developmental psychology, to wit, that there are few attempts to make sense of the accumulation of empirical data by means of basic models of the developmental process. There are, however, some very interesting developmental hypotheses and descriptions. For example, the hypothesis that stresses the relevance of redundant amodal information for the development of multisensory perception (Bahrick & Lickliter, 2012), the hypothesis that there is a regular developmental pattern of different abilities related to temporal information (Lewkowicz, 2000), or some attempts to identify the cognitive mechanisms responsible for the development of multisensory perception (Mareschal et al., 2012). As for the relationship between the development of multisensory perception and that of other domains, such as the social domain, some proposals reveal their connection. From the ecological perspective of perceptual learning and development (Gibson & Pick, 2000), the development of social interaction is seen as a result of an increase in the infant's capacity to differentiate perceptual information, insofar as it allows to perceive new affordances in parental behavior. In this regard, we have elsewhere reported (Martínez, 2019, see also Chap. 5 in this volume) an intimate link between the development of multisensory perception of temporal amodal information (detection, discrimination, and recognition of rhythm) and some changes in the infant's interactive behavior (regulation of his/her activation threshold by the adult, anticipation of the adult's behavior, turn-taking during joint actions) throughout the first year of life. This work – as proposed by Rivière (1986/2003) and Español

(2010a) – adopts a dialectical perspective of the relationship between the development of perceptual abilities and early intersubjective development. However, this shows only one side of the coin: how some developmental achievements in the multisensory perception of temporal amodal information allow us to understand the emergence of new modes of intersubjective contact. In this regard, the other side of the coin has yet to be revealed: how the genesis of the development of multisensory perception of temporal amodal information (hereinafter MPTAI) can be found in the interaction with adults during early intersubjective encounters.

In this chapter, we provide an explanation of the ontogenesis of MPTAI. To this end, the text will be organized as follows. First, we will define MPTAI and establish the coordinates that mark out the conceptual, empirical, and methodological framework wherein lies the explanation of its ontogenesis. Second, we will provide a prolegomenon to the explanation of the ontogenesis of MPTAI. In the third place, we will discuss the theory of perceptual learning and development proposed by Eleanor Gibson, which offers an explanation that partially accounts for the development of MPTAI. The theory has some shortcomings that are borne out and analyzed. This analysis allows us to glimpse a plausible way to overcome them. Fourth, in order to address unsettled problems in E. Gibson's account of the development of MPTAI, we appeal to a Vygotskian theoretical framework that conceptualizes development as the functional reorganization of different psychological systems. In this way, we look for a possible integration of developmental models with the aim of offering a solution to unsolved problems (Valsiner, 1998). Finally, we will introduce a model explaining the development of MPTAI, based on the reorganization of functional systems, as nested in intersubjective encounters between the infant and his/her caregivers.

6.2 Multisensory Perception of Temporal Amodal Information (MPTAI): Conceptual, Empirical, and Methodological Approaches for an Explanation of Its Ontogenesis

Multisensory perception consists of the ability to perceive in a unified and coherent manner the information that a single stimulus source provides simultaneously to two or more perceptual systems. This ability is also often referred to as intermodal perception or multimodal perception (Bahrick & Lickliter, 2009). In this chapter, however, we use each of these terms in a specific way, trying to avoid their use as mere synonyms. We reserve the term *multisensory perception* to refer to the ability to deal with stimuli from a single-source providing multiple sensory information. We use the term *intersensory perception* to refer to the ability to perceive relationships between information from a single source simultaneously available to different perceptual systems. To describe the amount of simultaneous information offered by a particular stimulus or event, we use the terms *unimodal information*, where

information is available for a single perceptual system, *bimodal information*, where information is available for two perceptual systems, and *multimodal information*, where information is available for three or more perceptual systems.

The development of multisensory perception faces the infant with an important developmental challenge: to be able to attend, as efficiently as possible, to multimodal information relevant to his/her needs and actions, while ignoring other less relevant stimulation (Bahrick & Lickliter, 2012).

6.2.1 Amodal and Modality-Specific Information

It was E. Gibson (1969) who, within the discussion on information transfer between different perceptual systems, noticed the important role of amodal information for perceptual development. Amodal – meaning *without* modality (Bahrick, 2009) – perception refers to the perception of information common or redundant to multiple perceptual systems. Amodal perception relates to *amodal information* whose defining characteristic is that it is extracted, indistinctly, through different perceptual systems. Unlike amodal information, *modality-specific information* can only be extracted by one perceptual system (e.g., color, smell, timbre of sound, etc.). For E. Gibson (1969), amodal information is inherent to the temporal and spatial dimensions of the environment. Properties of objects and events such as temporal synchrony, rhythm, rate, duration, intensity, and joint location are common to auditory, visual, and proprioceptive stimulation (Bahrick, 2009). According to Tees (1994), these amodal properties relate to space and time and involve points along a continuum (location), intervals within the continuum (duration), patterns of interval (form), rate of patterns (tempo), or change of rate (texture gradients). Thus, we should tell apart amodal spatial information from amodal temporal information. The joint location in space, or the edges and the angles of objects, for instance, are properties of spatial amodal information, whereas duration, rate, and rhythm are properties of temporal amodal information.

To bring the relevant issues into focus, we propose a synthetic definition[1] of the three kinds of amodal temporal information whose perception our model aims to explain. We should clarify that such a task is difficult and elusive due to the many attempts found in different disciplines (physics, musicology, music history, psychology) and also because the definitions themselves have been tinged by historical reasons (de la Jacquier et al., 2013). According to the Diccionario de la Real Academia Española (Dictionary of the Royal Spanish Academy) (2019), “duration”

[1] The definitions used in this chapter are influenced by research on music. This bias comes from the interdisciplinary dialogue between psychology of music and developmental psychology that guides the work of our research team (Español, 2014a). Although research on dance, somatics and music are equally favored in our team, the definitions shown below are more congenial with music research, partly due to the conceptual difficulties of the other two disciplines to break up movement into discrete units.

is defined as the action and effect of lasting; the time something lasts or that elapses between the beginning and the end of a given process. In this sense, it is possible to define duration in psychological terms as an ongoing psychological experience determined by the beginning and end of an event or occurrence. The term "rate" refers to the chronometric density of a sequence of events, understood as the number of events occurring in relation to a given time span (Pereira Ghiena et al., 2013). "Rhythm" is defined as the organization of temporal patterns of events or acts closely linked to temporal regularities; that is, to metrical structure. Metrical structure, in turn, is a psychological construct built on perceptual cues that are based on impulses, pulses, and relationships between pulses (Pereira Ghiena et al., 2013).

A particular case we should consider at length is multimodal temporal synchrony. We assume the argument proposed by Lewkowicz (2012) that establishes a functional distinction between amodal information and multimodal temporal synchrony. He considers duration, rate, and rhythm to be amodal attributes (amodal information) of the stimulus, since they can be specified independently across different perceptual systems and, as a result, can be perceived even in the absence of temporal synchrony cues. For example, even if the amodal information of speech sounds and orofacial movements, typical of linguistic communication, is not synchronously available, their duration can still be perceived. In contrast, temporal synchrony cannot be considered as an amodal attribute of stimuli because it cannot be specified independently through a single perceptual system. To perceive temporal synchrony, one must have access to simultaneous information available to at least two or more perceptual systems. Moreover, infants can perceive intersensory synchrony and fail to perceive amodal information. For example, an infant may perceive synchrony between the movements of the orofacial area and the sound of speech and miss the relationship between the duration of the two.

6.2.2 Different Perceptual Skills

The explanation of the development of multisensory perception of amodal temporal information involves accounting for a series of skills related to the perception of duration, rate, and rhythm. Basically, detection, discrimination, and recognition are the skills usually studied (Lewkowicz, 2000; Walker-Andrews, 1997). *Detection* simply indicates that an observer is sensitive and responds in some way to the information, i.e., when the organism is affected by a certain level or magnitude of information available to a perceptual system. *Discrimination* refers to the ability to differentiate between two or more objects, events, or between two aspects (i.e., information) of the same object or event. The ability to discriminate presupposes the ability to detect. Discrimination of what goes undetected is not possible. *Recognition,* which in turn presupposes detection and discrimination, refers to the ability to establish relations (in terms of equivalence or perceptual invariance) between two or more amodal characteristics of an object or event when they occur simultaneously.

There is some consensus in that each of these skills is studied by using different data collection techniques.

To study detection, we use techniques that identify changes in an organism's behavioral activity (voluntary movements of the hands, eyes, or body, among others) or involuntary activity (reflexes, heart rate, and pupil dilation, among others) when faced with the presence or manipulation of, mostly, the temporal amodal information magnitude presented in a multimodal manner. The *infant-controlled habituation technique* (Horowitz et al., 1972) is used to study discrimination. It involves the introduction of a pattern (duration, rate, or rhythmic) that models the amodal information that characterizes an object or a multimodal event through successive trials, called habituation trials. As the name suggests, they are used for habituating, for example, an infant to the temporal characteristics of an object or event. The criterion for habituation to a given stimulus is conventionally set when in the last three habituation trials the infant's visual attention toward the stimulus falls 50% below that during the first three trials, or when the infant has finished all habituation trials. This is followed by the test phase, when the object or event is shown a second time with the amodal characteristics to which the infant became habituated, followed by the presentation of an object or event whose amodal features differ from those of the stimulus the infant was habituated to. We infer the infant discriminates temporal amodal information according to its dynamic modeling (in terms of duration, rate, or rhythm) when a longer attention time, usually visual, becomes evident when faced with the new information presented, compared to the information the infant was habituated to.

Recognition is studied using the *intermodal preference technique* (Spelke, 1976). It consists of the simultaneous presentation of information, generally visual, corresponding to two objects or events (using two monitors, located side by side) and that of synchronous and concordant auditory information (using a loudspeaker located between the two monitors) with only one of the visual objects or events.

Concordance and synchrony between the multimodal information available to the visual and auditory perceptual systems lies in some of the dynamic qualities (duration, rate, or rhythm) of the amodal information. In this sense, amodal information characterizing one of the events or objects displayed for the visual perceptual system exhibits the same perceptual invariance (intersensory relationship) as the amodal information characterizing the information available for the auditory perceptual system.

We calculate the proportion of total looking time (PTLT) for each infant by dividing the visual fixation time on visual information congruent with auditory information by the sum of the visual fixation time on visual information corresponding to congruent plus incongruent events. A statistically significant difference favoring PTLT on visual information congruent with auditory information indicates the recognition of the intersensory relationship or perceptual invariance.

6.2.3 Development of Perceptual Skills

The ontogenesis of MPTAI involves the emergence of different skills (detection, discrimination, and recognition) associated with different properties of temporal amodal information (duration, rate, and rhythm). Lewkowicz (1994, 2000) raised two relevant issues about the development of perception of amodal information. First, he noted the same developmental sequence for the three skills described above, regardless of the kind of information involved. Empirical evidence suggests that, whether for duration, rate or rhythm, detection and discrimination emerge earlier than intersensory recognition. Second, he described a potential developmental sequence for the recognition of intersensory relationships based on different types of amodal temporal information. At first, infants recognize intersensory relations based on duration to later recognize them according to rate. Regarding rhythm, no evidence on its recognition was available at that time.

Since it involves different abilities and types of information, it is quite obvious that the development of multisensory perception of amodal temporal information is a complex process. Recently, based on a review of available empirical data and our own findings, we proposed a detailed developmental sequence of multisensory perception of rhythm (M. Martínez et al., 2018, M. Martínez, 2019, see Chap.5). Briefly stated, during the gestational period (Provasi et al., 2014) and for the first few months of life (Lewkowicz, 2000), infants are sensitive (they detect variation) to different rhythmic patterns. Between the ages of 4 and 10 months, infants discriminate between two rhythmic patterns presented bimodally (sound and movement) (Pickens & Bahrick, 1995; Lewkowicz & Marcovitch, 2006); at 7 months they can identify the same rhythmic pattern when first presented, for example, to the visual perceptual system and subsequently to the auditory system (Allen et al., 1977); from 8 months, in addition to discriminating bimodally presented rhythmic patterns (visual and auditory), infants also discriminate between two rhythmic patterns displayed unimodally, i.e., visually (Bahrick & Lickliter, 2004). Starting at 10 months, infants discriminate between two bimodal rhythmic patterns even when visual and auditory information are presented simultaneously but not synchronously (Lewkowicz, 2003); at this same age, infants recognize rhythm-based intersensory relationships (Martínez et al., 2018).

A potential difficulty in describing the emergence of different skills lies in the limited available data concerning what happens once the ability is acquired. For example, Lewkowicz (1986) conducted a study with infants aged 3, 6, and 8 months. According to his data, 6- and 8-month-old infants recognize intersensory relations based on duration. However, the findings failed to demonstrate potential differences in recognition at 6 and 8 months. Addressing these issues is highly relevant, since several authors (Bahrick, 2010; Gibson, 1969) argue that an explanation of the ontogenesis of multisensory perception should account for the increased effectiveness of the skill as well. Multisensory perception not only involves changes in different kinds of skills and information, but also changes in the optimization of the ability to extract information. We found empirical evidence suggesting that 10-month-old

infants are more efficient than 4-month-old infants in recognizing duration-based intersensory relationships, because they need less time to extract such information from the environment (M. Martínez et al., 2022).

In summary, an explanation of the ontogenesis of multisensory perception of amodal temporal information involves accounting both for the emergence and for the increased efficiency of different abilities (detection, discrimination, and recognition) linked to different types of information (duration, rate, and rhythm). According to E. Gibson (1966), differences in efficiency is one possible change in perceptual development. A further and currently well-studied change is perceptual narrowing (Bremner et al., 2012). These are not necessarily the only ones, so the door remains open for additional changes to occur and for future conceptualizations and empirical studies to accrue.

6.3 Setting the Ground for the Explanation of the Genesis of MPTAI

The enterprise of providing an explanation of the ontogenesis of MPTAI requires that a number of issues be delimited and framed. Failure to properly address these issues might undermine the whole enterprise.

6.3.1 First Issue: The Prominent Role of Multisensory Perception of Temporal Amodal Information

The first issue to be addressed is to duly account for the reason our explanation only focuses on the development of multisensory perception of temporary amodal information, leaving aside the development of other capacities, for example, the one concerning the perception of intersensory relations based on modality-specific information. Basically, we base our choice on four reasons intimately linked to the general development of perception and to early intersubjective contact.

First, as Lewkowicz (2012) points out, the temporal dimension of our everyday experience is an unavoidable and fundamental part of our perceptual and cognitive existence: the temporal flow of stimulation provides a series of perceptual cues that observers can use to detect the coherence, the overall structure, and even the hidden meanings of multimodal events.

Second, during early intersubjective encounters, whether dyadic or triadic, infants progressively become active participants in multimodal performances where they observe adult movements, hear adult sounds, feel adult touches on their bodies, and experience adult causal agency when moved by them. Encounters with adults are a multisensory experience for infants. During the interaction, adults simultaneously provide information to the infant's various perceptual systems. Such

information is temporally modeled with respect to duration (Stern, 2002), rhythm (Koester et al., 1989), synchrony between sound-gesture and kinetic-gesture (Martínez, 2014), and the repetition-variation form (Español & Shifres, 2015– see also Chap. 2). Temporal amodal information is deliberately arranged by adults to capture and sustain the infant's attention during early intersubjective exchanges (Martínez, 2019).

Third, the discrimination of bimodal and unimodal rhythmic patterns allows the infant to segment and anticipate adult behavior, while the recognition of intresensory relationships based on rhythm and duration are closely linked to adult behaviors directly oriented to intersubjective contact, such as affect attunement (Martínez, 2019). Through affect attunement, the adult expresses and shares affective states with the infant (Stern, 1985). On the other hand, infants respond in a different way to affect attunements during the last semester of the first year of life (Bordoni et al., 2016). Affect attunement is a frequently used resource, oftentimes more than imitation, to establish adult-infant intersubjective contact (see Chap. 4).

Finally, the fourth reason why we decided to concentrate on the perception of intersensory relations based on amodal information, while explicitly leaving aside the development of the perception of relations based on modality-specific information, is that intersensory relations based on the latter are arbitrary (Walker-Andrews, 1994) and therefore must be learned (Bahrick, 2001). Consequently, we can assume that the development of recognition of intersensory relations based on modality-specific information involves different processes, most likely inherent to information integration (Lewkowicz, 2012). Furthermore, plenty of evidence indicates that the development of the perception of intersensory relations based on amodal information precedes the recognition of intersensory relations based on modality-specific information (Bahrick, 1992, 1994, 2001, 2010). Thus, inasmuch as intersensory relations based on amodal information lend coherence to the temporal flow of information (Lewkowicz, 2012), they should be considered the cradle for subsequent recognition of intersensory relations based on modality-specific information (Bahrick, 2004).

6.3.2 Second Issue: Opting for a "Differentiation" Perspective

A second issue we must address is the need to justify our position on whether the development of multisensory perception proceeds through information integration or differentiation. Each of these approaches originates in a particular way of understanding perception and how it should be studied.

On the one hand, psychological explanations of perception are usually tinged with a certain visuocentrism (Travieso & Blanco, 2000). This means that the study of perception is mostly focused on the visual system and holds the assumption that the functioning of other systems is subject to the same rules, at least at a cognitive level (Dember & Warm, 1979). On the other hand, the way perceptual systems are classified has a direct impact on the study of multisensory perception. Traditionally,

different types of receptor cells, afferent pathways, and primary cortical areas are used to identify those systems (García-Albea, 1993). Visuocentrism and the biological distinction of the senses have a direct impact on how to explain the genesis of the phenomenological experience of unified information coming through the different perceptual systems, and lead to a particular way of approaching the problem of multisensory perception and its development. Specifically, the development of perception is assumed to be an integrative process. The senses, or rather the perceptual systems, are separate at birth and the infant gradually learns to integrate sensory information from different sensory modalities. Before integration is achieved, the infant perceives a scattered stream of light, sound, tactile, and proprioceptive impressions. According to Bahrick (2009), Brich and Lefford (1963) and Piaget (1977) are some of the most relevant authors who embraced this perspective.

As already mentioned, the essential feature of amodal temporal information is that it can be perceived indistinctly by different perceptual systems. This feature poses a problem to the integration perspective: amodal temporal information does not need to be integrated, at least from a logical point of view. In this sense, its development must be explained from a different perspective. For this reason, we believe that the genesis of MPTAI must be approached from a different angle from integration, one that assumes that the infant's developmental task is to distinguish the set of invariants characterizing the information arriving simultaneously to the different perceptual systems. This approach may be called the "differentiation perspective"' and was assumed by authors like Bower (1974), E. Gibson (1969), and Werner (1948). According to Lewkowicz (2000) and Bahrick and Hollich (2017), empirical data support the differentiation perspective with respect to amodal information perception.

6.3.3 Third Issue: Perception/Action and Its Genesis

A third issue is to avoid committing ourselves to a perspective that hinders or limits ontogenetic explanation, as occurs when the perception/action dualism is explicitly or implicitly assumed. According to Aivar and Fernández (2006), most psychological theories of perception take reality for granted while at the same time define sensation as the contact between this given reality and people, conceiving it as a mere physico-mechanical reception process. From this perspective, movement is seen in active terms and sensation in passive terms. Movement and activity are considered identical terms and opposed to sensation (Aivar & Fernández, 2006). This is probably the major fallacy of explanations elaborated under the classical cognitive perspective, such as the one proposed by Mareschal et al. (2012).

Other approaches, in contrast, avoid the problems associated with dualism and the passive view of perception. Blanco (2002) cites Travieso's work on haptic perception (Travieso, 2000; Travieso & Blanco, 2000) as an example of a psychological explanation that circumvents these difficulties. The ecological theory of perception (Gibson, 1966, 1979), the enactive perspective (Varela et al., 1991),

together with the idea of functional systems (Luria, 1973; Vygotsky, 1930/1999) offer a tempered space, to use Blanco's expression, where we find it possible to give an account of the ontogenesis of MPTAI. Similarly, very recently, Read and Szokolszky (2018) and Szokolszky and Read (2018) proposed a similar framework referred to as *ecological developmental psychology*. Our own proposal fits quite well with these views, albeit in a somewhat nuanced way. While the general perceptual framework is the ecological theory proposed by J. Gibson, we will also assume other theoretical commitments from the theory of perceptual learning and development proposed by E. Gibson (1969).

Finally, one last issue we would like to address– essential to any attempt at an ontogenetic explanation– is the need to include the temporal dimension to explain how behavior unfolds over time, and how the mechanisms that make it possible are transformed (Rosa et al., 1993). Therefore, we must establish precisely which are the emerging perceptual abilities and which mechanisms or processes make them possible. A comprehensive theory of the development of MPTAI should account for the emergence of different skills (detection, discrimination, and recognition) as they are linked to different kinds of amodal temporal information (duration, rate, and rhythm). Regarding the mechanisms or processes underlying changes over time, the theory of perceptual development developed by E. Gibson satisfies this requirement.

6.4 The Ecological Theory of Perception and Perceptual Learning

Eleanor Gibson developed a theory of perceptual learning. The first systematization appeared in her work *Principles of Perceptual Learning and Development* (Gibson, 1969). In her last book, co-authored with Ann Pick, *An Ecological Approach to Perceptual Learning and Development* (Gibson & Pick, 2000) she introduced her latest formulation. Her theory is consistent with the ecological theory of perception proposed by James Gibson (Adolph & Kretch, 2015). It assumes certain principles while postulating two central concepts to explain perceptual learning: differentiation and exploratory activity. According to Gibson and Pick (2000), the elements from the ecological theory of perception underlying the explanation of perceptual learning are a general principle (reciprocity between the environment and the perceiver) and three fundamental concepts (affordance, information, and pickup of information).

Reciprocity between the environment and the perceiver. Perceiver and environment constitute an interactive system. The environment provides opportunities and resources for action, as well as information determining what is perceived while guiding the action. In turn, action brings about consequences that provide the perceiver with additional information. Animals generate information about the environment and adapt their action to it; reciprocally, the environment provides the animal with opportunities and resources for action (E. Gibson & Pick, 2000).

Affordances are possibilities for action, what the environment offers the animal (J. Gibson, 1979). They depend on the fit between the animal's bodily capabilities and the environment's physical properties. Affordances are environmental attributes related to the animal's capacity for using them. According to E. Gibson:

> Babies have a great deal to learn about – everything that the world has to offer – and perceptual learning is their way of discovering what particular things and people afford for them, where things and people are in relation to themselves, what is happening, what characterizes their permanent surrounding, and what they can do (Gibson & Pick, 2000, p. 21).

For E. Gibson, the outcome of perceptual learning are new affordances and new characteristics or physical properties of events. But, above all, what animals learn to perceive are affordances for action. The concept of affordance implies a reciprocity between perception and action. The function of perception is to gather information about objects, events and places; and find out a potential use for it. E. Gibson argues that all actions (walking, cooking, writing, driving a car) are guided by perceptual information. Reciprocally, action reveals the animal information about the environment, itself, and what it might do next. Although E. Gibson distinguished between exploratory action (information seeking) and executive action (control of environmental consequences), both types of action generate perceptual information, and are driven by perceptual information (Adolph & Kretch, 2015).

Information. The ecological approach differs from traditional theories of perception in the way perceptual information is conceptualized (E. Gibson & Pick, 2000). Perceiving does not involve structuring the impoverished sensory stimulation that reaches the sensory receptors, but rather detecting the structure existing in the environment. We do not perceive the unfolding of environmental energy or information; on the contrary, by extracting the organization or order from the environment (ambient arrays) we perceive things in the world that give structure to information and are specified by information. Within perceptual learning theory, the term information denotes the structured distribution of the environment's organization energy specifying the events or aspects of the environmental events. The sources of information are events, objects, and the layout of the environment's surfaces. In Sect. 6.4.4, we will review the concept of information more closely.

Pickup of information. E. Gibson assumes that animals, including humans, actively seek information to guide their actions and learn about their environment. Information is not something that comes to the organism, but something the organism extracts from the environment. Perception guides action, action makes information available to perception (E. Gibson & Pick, 2000). In this sense, to perceive is to act; and action leads to information extraction.

6.4.1 The Explanation of Perceptual Learning

More than thirty years passed since the first formulation of E. Gibson's theory of learning and perceptual development and her last presentation. In her 1969 book she described her theory in detail and launched a new field of research, intimately linked to the role of experience and learning in the changes in perceptual skills. As time went on and new research methods and consequent findings became available, mainly related to infant perception, E. Gibson questioned some of her initial assumptions and expanded her research domain. Specifically, she criticized her initial proposal for its inability to capture how infants learn to detect perceptual information to guide their action adaptively as their perceptual-motor systems develop (Adolph & Kretch, 2015).

In her first formulation, E. Gibson defined perceptual learning

> as an increase in the ability of an organism to get information from its environment, as a result of practice with the array of stimulation provide by the environment. This definition implies that there are potentials variables of stimuli which are not differentiated within the mass of impinging simulation, but which may be, given the proper conditions of exposure and practice (E. Gibson, 1969, p. 77).

According to this definition, there are potential variables within the undifferentiated mass of stimuli that impacts on the organism, but given the right exposure and practice conditions, these variables may be differentiated. This means that a change has been caused in the stimuli to which the organism responds. This change does not imply the acquisition of a new response, or its substitution for another response to a stimulus that was previously responded in a different way, but rather it involves responding discriminately to a previously unanswered stimulus variable. Thus, the criterion of perceptual learning results from an increase in specificity. What is learned can be described as the detection of distinctive properties, patterns, and characteristics.

The 2000 version of the theory (E. Gibson & Pick, 2000) is a theory of development, rather than perceptual learning. Perceptual learning involves an increased ability to detect information that specifies affordances, events, and their distinctive features. Perceptual learning can be defined as the way that an organism discovers invariances in objects, events, and environmental designs. The reduction of a wide variety of information to minimal and optimal information specifies an event's *affordance*. The book written by E. Gibson and Pick (2000) devotes three chapters to discuss how infants learn about (a) communication, (b) interaction with objects, and (c) locomotion and spatial design. In reviewing these topics, the authors emphasized the role of perceptual learning, showing how it drives and constrains perception and action systems; and consequently, how it changes the developing infant's relationships with the environment. In this regard, we believe the latter formulation of E. Gibson's theory can be understood as a general theory of development, over and above a theory of perceptual development. On the other hand, while there is continuity with how the explanation of perceptual learning is formulated, the more recent version underscores the relevance of learning with respect to the concept of

affordance, leading to an expansion of the theory that, to our understanding, makes it a powerful tool to understand ontogeny beyond perceptual development. Likewise, two concepts that were present in the first version and are more deeply developed in the newer one – i.e., differentiation and exploratory activity – are extremely useful to explain the development of multisensory perception of amodal temporal information.

6.4.2 Differentiation

E. Gibson elaborated her explanation of perceptual learning focusing on differentiation and increased specificity, thus departing from traditional theories of perceptual learning (Adolph & Kretch, 2015). Gibson and Gibson (1955) defined perceptual learning as a process of differentiation, selection, and extraction of previously undetected information in the environment. They viewed learning as a process of differentiation of previously vague or diffuse impressions, not an enrichment process of previously poor sensations.

The task of the perceiver is to extract the information needed to specify the affordances in the environment; identifying such information is the result of perceptual learning. Perceptual differentiation allows an organism that receives a large amount of simultaneous information to reduce it to what is needed to specify the affordance of an object or event. Differentiation occurs both at a global and a local level of the object or event, i.e., with respect to the features that specify the object or event (Goldstone, 1998). For amodal temporal information, for example, global differentiation is linked to event duration, while the differentiation of distinctive aspects of events is related to the rate or the rhythmic structure.

E. Gibson (1969) claimed that discriminative learning is linked to certain characteristics of the experience undergone by the organism on the basis of which differentiation occurs, thereby leading to perceptual learning. In other words, it is linked to repeated experiences that enable the organism to respond to variables of physical stimulation previously ignored. To this end, she considered two essential characteristics.

Emphasis on distinguishing characteristics. Discrimination is facilitated when differences between distinctive features are emphasized during the learning experience.

Contrast. Learning differential properties is facilitated when examples are provided during the learning experience that enable to contrast dimensions in such a way as to help define and isolate the specific property of the critical variable that must be differentiated. The stepwise contrast is particularly helpful to isolate a continuous relational property. That is, the gradual difference between two properties of a stimulus is more beneficial to isolating a property than abrupt differences.

6.4.3 *Exploratory Activity*

According to E. Gibson, exploratory activity is an integral part of the perceptual learning process. Animals seek information about the relationships between the self and the surrounding environment (E. Gibson, 1997). Evolution provides animals with perception-action systems which furnish them with the means to discover what the world is all about. The development of these systems provides the motivation to explore an accessible environment, to act upon it, and to extend explorations as the organism's capabilities increase (Adolph & Kretch, 2015).

Exploratory activity includes behaviors such as scanning, focusing, groping, articulating, licking, listening, shaking, smelling and manipulating, and searching for information by using the most appropriate actions and body parts for an efficient search (E. Gibson, 1970). The exploratory activity is performed on all types of available information: in this sense, it is multisensory. Although every movement of the organism produces perceptual information, E. Gibson (1988) believed that focused and deliberate exploratory behavior is critical for perceptual learning. In her view, "A sequence of acts termed exploratory will have some outcome and will not be random. It will have a perceptual aspect, a motor aspect, and knowledge-gathering aspect." (E. Gibson, 1988, p. 5). Realizing the consequences of one's exploratory actions leads to learning.

If perceptual learning is preeminently a process of differentiation, the organism should be able to select between different simultaneous kinds of information and different actions. For E. Gibson, the wide variety of information is ensured by the youthfulness of the offspring of different species that are highly motivated to use their perceptual systems and repertoire of actions to explore themselves and their surroundings. In this respect, the immaturity of the offspring ensures the development of exploratory activity (E. Gibson & Pick, 2000).

The information specifies the design of the environment and the events surrounding the organism. This invariant information must be extracted through the appropriate perceptual system. Infants are natural explorers who constantly seek to engage with their environment through all available means (E. Gibson & Pick, 2000).

To E. Gibson, exploratory activity is a perception-action sequence with its own consequences. Exploration brings new information of two kinds: about action-related changes in the world and, at the same time, about what the organism is doing. Extracting information that specifies affordances is not an accidental activity. E. Gibson and Pick (2000) emphasize that when observing an infant, he/she is clearly affected by his/her own active gaze, auditory awareness, and movements of the limbs and head. Exploratory activity ensures changes in the infant's interaction with the information around him/her and in the simultaneous changes in his own body, which, as E. Gibson (1988) argues, are essential for perceptual learning.

The development of postural control (head, trunk, and eventually the whole body) in infants creates new opportunities for exploration and perceptual learning (E. Gibson, 1988). During the first weeks of life, infants gain control over their eyes and head, which allows an active exploration of the visual world. During this period,

infants learn about objects, people, and events by actively directing their gaze to them. Later, infants are able to sit independently. This sets their hands free and leads to improve their manual skills, such as the ability to explore objects with their hands, to coordinate visual inspection with manipulative actions, and to alternate between oral and visual object exploration. In this period, they learn about the physical properties and affordances of objects. Toward the end of the first year of life, when infants begin to crawl and walk, they discover an expanded world at their disposal. During this period, they use locomotor exploration to learn about the properties and affordances of surfaces, places, and spatial design. In this regard, one may recall the classic visual cliff experiment conducted by Walk and Gibson (1961). Perceptual-motor development is the main agent of early perceptual learning. The onset of a new action system, such as walking, enables a new exploratory activity that generates information about new bodily capacities (as information extraction systems) and about the properties of the environment involved in novel actions performed by the infant on it (E. Gibson, 1988). Exercising particular exploratory activities results in the optimization of information retrieval, and also in the development of more complex forms of coordination (Adolph & Kretch, 2015).

6.4.4 Scope and Limits of Perceptual Learning Theory

We believe the theory of perceptual learning and development proposed by E. Gibson is a relatively plausible framework to account for the development of MPTAI. We deem it *plausible* because by assuming the general framework proposed by J. Gibson, the theory adequately addresses the perception-action relationship. Also, because by proposing differentiation and exploratory activity as responsible for perceptual learning, the theory allows to account for the changes resulting from the extraction of information. On the other hand, we think this framework is only *relatively* plausible because the theory exhibits certain weaknesses. In particular, due to its commitment to J. Gibson's ecological perspective, it assumes a direct realism that fails to circumvent the perception-action dualism (Aivar et al., 2002).

Regarding direct realism, Aivar and Travieso (2009) point out that the problem that the ecological theory of J. Gibson, as well as E. Gibson's (1969) perceptual learning theory and its more contemporary derivations such as O'Regan and Noë's (2001) theory, fail to solve lies in the assumption that information is pre-existent to the subject who extracts it from the environment. That is, information is sitting *there* permanently, waiting to be extracted by the organism. Ultimately, the question is how to solve the problem of the organization of information, or the independent characterization of the distal stimulus. García-Albea (1993) lists various ways of approaching the problem: in classical cognitive psychology, perceptual symbolic mechanisms interpret the stimulus according to its physical description; in *Gestalt* theory, the structure that shapes the information is within the mind in the form of

laws of perception; as for the ecological theory of perception, the structure that shapes the information is in the information itself: the laws of ecological optics.

We propose a fourth possibility. By focusing on amodal temporal information (duration, rate, rhythm), we can reframe the problem of realism by rejecting the notion that information is always *there*. Information is there as a consequence of and during interaction. Amodal information becomes organized as a result of the infant's interaction with his or her caregivers. Adults provide information in a redundant, temporally organized way, through a repetition-variation form or by means of contrasts that come about as a result of the caregiver's engagement in the interaction with the infant. If the ontogenesis of temporal amodal information perception abilities is rooted in adult-infant intersubjective contact, we believe that temporal amodal multimodal information – duration, rate, or rhythm – gets organizes as a consequence of the exchanges between the developing organism and another adult organism. Information does not exist prior to the organism.

According to J. Gibson (1979), the term information refers to the specification of the observer's environment. The organization of information from the wave of energy (luminous, mechanical, or chemical) surrounding us is invariant. This information is what J. Gibson (1979) considers to be out there to be extracted. Contrary to what would happen in the physical world, we assume that the organization of temporal amodal information is not there waiting to be extracted. Conversely, the organization of amodal temporal information emerges during intersubjective contact, as a function of the adult-infant mutual involvement. The organization of amodal temporal information in adult performance depends on the participation of the infant. We will later see that this participation is not mere social contact; the infant's participation is the set of actions he/she performs in order to extract information. The actions performed by the infant simultaneously serve a double function: extracting the amodal temporal information and contributing to its dynamic organization in time.

In our view, the problem posed by direct realism might begin to be solved if we understand the ontogenesis of MPTAI as framed within intersubjective contact. Similarly, we may trace the phylogenesis of the role of amodal information in the psychological development of human creatures to an intersubjective context. Español (2010b) perfectly synthesizes this very idea. The precocity of birth in humans affected the social world of infants, since their immaturity inevitably led to dependence and to an ever-present human parental scaffolding from birth. Human offspring's long period of immaturity generated a selective pressure for mechanisms that could guarantee psychological contact, which would in turn make intersubjective experiences possible. Such selective pressure resulted in a co-evolution of universal predispositions in both parental and filial populations that operate early in ontogeny and are controlled by non-conscious subsystems of behavioral regulation. The bonding modality with the conspecific changed radically: "Selective pressure led to filial behaviors present in other primates – such as facial expressions, gestures, movements and sounds – elaborated or treated in a special way, through dynamic, rhythmic and transmodal shaping" (Español 2010b, p. 50). We can therefore think of an alternative idea to the one that holds that information is *there*. We

claim that infants acquire psychological experience of temporary amodal information based on their intersubjective encounters with adults.

A second problem faced by E. Gibson's theory concerns the ontogenesis of exploratory behavior. The theory of perceptual learning and development grants a natural character, a natural genesis, to the development of exploratory activity, the cause of its development being the maturation and progressive postural control of the organism (E. Gibson, 1988). E. Gibson believes that exploratory activity is inherent to perceptual learning. The organism has perception-action systems for this activity. The development of these systems ensures the necessary motivation to explore the environment. The variety of available information is ensured by the offspring's youth, and its natural motivation to use its perceptual systems. Identifying the consequences of one's own exploratory actions leads to learning and, in this sense, perceptual-motor development is the vehicle for early perceptual learning. Changes in perceptual skills are ultimately due to the development of voluntary control of actions on the environment, which depends on the development of postural control (E. Gibson & Pick, 2000).

We believe problems of direct realism and of the natural character of the genesis of exploratory activity are adequately solved if we conceive ontogenesis within the framework of intersubjective contact and from the perspective of the development of functional systems. An account of the development of MPTAI can be given on the basis of the explanation offered by E. Gibson (E. Gibson, 1969, E. Gibson & Pick, 2000) on differentiation-based perceptual learning and development. Yet this explanation ought to be reformulated from a Vygotskian perspective, especially with regard to the ontogenesis of exploratory activity. One way is by conceiving the development of MPTAI in terms of the construction of functional systems for action (Rosa et al., 1993).

6.5 The Development of Multisensory Perception as a Development of Functional Systems

Turning to the Vygostkian framework to account for MPTAI's development can seem somewhat puzzling. The theory of the Russian psychologist is usually linked to the development of higher psychological processes, closely connected with the processes of semiotic mediation. This is most likely due to the particular historical process that gave rise to standard Vygotskian thought in Spanish-speaking countries, an historical process that set Vygotsky as a canon and imposed the task of defining the central topics of his work (García, 2016). This canon may have led to a biased interpretation of the development of perceptual processes as elementary processes. Under this canon, the theory assumes that elementary psychological processes develop under the natural law of development according to maturation and independently of social interaction. In this manner, we might believe that Vygotsky barely referred to elementary processes and did so on grounds of a reflexological

model where the word "elementary" suggests that these functions are indeed elementary, a notion that is hard to uphold nowadays (Rivère, 1999/2003). With this view of perceptual processes in mind, it is certainly puzzling to resort to socio-historical-cultural theory to try to account for the genesis of the MPTAI.

In different works, Vygotsky (1978, 1930/1999, 1932/1987) approached the development of perception by providing ideas and conceptual tools that encourage us to go beyond the canon, thus bringing us closer to the problem of the development of MPTAI. Vygotsky directly addressed the issue of the integration of information from different perceptual systems. He started his analysis by identifying the limits of the associationist and structuralist theories available at the time.

Regarding associationism, Vygotsky noted that although his representatives agreed "that separate sensations are present from the beginning, researchers differed in their opinion as to which month of the first year of life is associated with the development or emergence of these complex perceptions" (Vygotsky, 1932/1987, p. 291). Concerning structuralism, he said: "Structural psychology fell into a dangerous trap in its treatment of the child's perception […] The structural character of perception is primal; it emerges at the outset. It is not the product of an extended development" (Vigotsky, 1932/1987, p. 291). His analysis showed how both approaches failed to account for the genesis of human perception. His conclusion was:

> Whether we consider orthoscopic perception, meaningful perception, or the connections between perception and speech, we find a single fact of extraordinary theoretical importance. Specifically, at each stage in the child's development we observe changes in interfunctional connections and relationships. In the child's development a connection arises between the function of perception and the functions of eidetic memory. As a consequence, a new unified whole arises, a whole within which perception acts as an internal constituent (Vygotsky, 1932/1987, p. 299).

In the above statements, we appreciate that Vygotsky proposes to account for the development of perception from the perspective of the development of functional psychological systems (Vygotsky, 1930/1997, 1930/1999), whose genesis is explained on the basis of the development of functional connections and interconnections. We believe the development of MPTAI can also be approached from the point of view of a functional psychological system that develops from new functional connections.

Vygotsky defines psychological systems as the emergence of new and changing relationships between different psychological functions (Vygotsky 1930/1997). Development – ontogenesis – not only involves the internal reorganization and refinement of isolated psychological functions (perception, memory, etc.); on the contrary, it implies that connections between functions are radically altered (Vygotsky 1930/1999). Missing connections between functions in the initial stages of development results in "new psychological systems arise that unite in complex cooperation a number of separate elementary functions" (Vygotsky 1930/1999, p. 61). According to him, the functional reorganization of psychological functions giving rise to specifically human psychological processes crucially depends on the internalization of semiotic instruments, especially language. Words or signs bestow

an indirect, mediated character to perception as a higher-level psychological process (Vygotsky 1930/1999). The internalization of semiotic instruments poses a challenge to our attempt to account for the development of MPTAI. In particular, two issues need to be settled: one is to determine the influence of social interaction on the development of elementary psychological processes, and the other, to establish how the internalization of semiotic instruments gives rise to MPTAI.

Regarding the influence of social interaction on elementary psychological processes, Van der Veer and Van Ijzendoorn (1985) highlight, on the one hand, studies (carried out by Zinchenko) revealing the fundamental role of continuous interaction with a culturally determined environment for the development of visual orientation and, on the other hand, the influence of pre-verbal interaction for the development of low-level or natural processes. Wertsch (1985), in turn, proposed that Vygotsky's thesis on the independence between the cultural and natural lines of development during the initial phases of ontogeny is too simple in light of our current knowledge on infant cognitive development. He also recalls how, in this regard, Leontiev and Luria stated that "after all, even in children at the very earliest ages mental processes are being formed under the influence of verbal social interaction with the adults around them. Consequently these mental processes are not 'natural'" (1960, p. 7, cited in Wertsch, 1985).

In this sense, referring to some works of Vygotskian persuasion carried out during the 1970s, Wertsch argued that it seems to be more appropriate to think that the cultural and natural lines are interrelated from the early stages of development. Furthermore, Wertsch (1985) emphasized that Piaget's work on the ontogenesis of sensorimotor intelligence demonstrates that Vygotsky's notion of natural development must be revised to incorporate other factors beyond maturation. He specifically stated that:

> Specifically, it would mean that the natural line of development could not be explained solely on the basis of organic maturation. Instead, a large set of explanatory principles would have to be employed, including those proposed by Piaget for the development of sensorimotor intelligence (Wertsch, 1985, p. 47).

Regarding the second problem, Vygotsky's interest in social processes led him to examine the representational systems that are required to participate in these processes, hence the emphasis on speech internalization (Wertsch, 1985). According to Vygotsky, internalization can only be fully understood by resorting to semiotic mediation (Martí, 1996). This requirement limits our possibility of understanding the development of MPTAI from a Vygotskian perspective based on the internalization of semiotic instruments of the infant's external control of action.

This is only partially true if we think that the natural line of development is not affected by the child's participation in social practices and if we restrict the concept of semiotic mediation to language. However, we have seen that, on the one hand, perceptual development can be thought of as culture-bound, and on the other, semiotic mediation far exceeds linguistic mediation (Español, 2003; Español, 2004; Rivière & Español, 2003; Rivière & Sotillo, 1999/2003; Rodríguez & Español, 2016). Nevertheless, the problem will not be fully settled until we specify the

mechanisms of semiotic mediation at work in the extraction of temporal amodal information.

The differentiation established by Rosa et al. (1993) between natural, combined, and artificial systems offers a way to explain the ontogenesis of the MPTAI from the perspective of functional systems. Natural systems are defined as the natural resources that biology offers to psychological functioning, and the unfold through the organism's biological maturation. Combined systems result from the combination of natural systems following the organism's activity in relation to its environment. The third kind of psychological systems, artificial or culture-bound, have a similar structure to combined systems, but differ from them in that the interfunctionality of combined systems depends on mediation mechanisms, tools, or instruments, among which linguistic signs are the most important.

To address the question of internalization, we suggest that the genesis of MPTAI be interpreted as the progressive development of combined systems. Rosa et al. argue that the combination of natural systems emerges as a consequence of the organism's activity with the environment. They result "from an apprenticeship through which the organism learns action patterns adjusted to the needs of its relationship with the environment" (Rosa et al., 1993, p. 325). MPTAI development would result from learning actions allowing the organism to extract information from the environment. Learning would allow the organism to discriminate between different kinds of information from the environment, resulting in new affordances. In this sense, adjusted action patterns would be those that the organism performs to extract information. Following Piaget, internalization is thus conceived of as the construction of schemas reflecting regularities of infant action based on external mechanisms of behavioral control. Wertsch (1985) stresses that internalization is a compatible space between Piagetian and Vygotskian postulates.

Perceptual development depends on the internalization of regularities of infant actions that are used to extract information. These actions are performed with the infant's perceptual systems and generate different sensory-motor contingencies (O'Regan & Noë, 2001). Actions enabling the extraction of temporal amodal information from adult performance initially occur in an inter-psychological fashion (through the mediation of adult performance) and are later carried out intra-psychically. According to Vygotsky (1932/1998), the infant's extreme dependence on adults, which goes beyond biological needs, "creates a completely unique character of the child's relations to reality (and to himself): these relations are always mediated by others, and are always refracted through a prism of relations with another person" (p. 216). The development of interfunctional relationships between natural systems follows the internalization of action regularities carried out using perceptual systems. These regularities reflect the cultural mode in which adult behavior is organized to attract and sustain the infant's attention toward interaction. It is also possible to identify how during intersubjective contact, adults regulate the baby's exploratory *activity* by using cultural forms of managing infant participation during intersubjective encounters. This aspect will be further elaborated in Sect. 6.6.2.

We believe that it is possible and useful to reconcile the Vygotskian and Gibsonian approaches to explain MPTAI development given their common conception of the perception-action relationship. We have already mentioned that according to ecological theory, perception entails action. Vygotsky, in turn, argues that in primary reflexes and in the simplest reactions perception and movement merge, "convincingly shows that both these parts are inseparable points of one dynamic whole, one psychophysiological process" (Vygotsky 1930/1999, pp. 30–31). In this sense, both approaches hold the same epistemological commitment to the relationship between perception and action.

On the other hand, the Vygotskian framework also copes with the problem of realism in E. Gibson's theory, while being compatible with her notion of differentiation (based on contrast and an emphasis on distinctive features) as a mechanism of perceptual learning. The problem of E. Gibson's theory lies precisely in its failure to explain the genesis of action, and its explanation of the development of exploratory activity from a maturationist or naturalistic standpoint.

6.6 A Theory of the Ontogenesis of MPTAI

6.6.1 From Natural Systems to Combined Systems

In this section, we will define natural systems whose functional reorganization results in the emergence of combined systems on the basis of which MPTAI development is explained. Our starting point is the presence of two natural systems: the detection of amodal information and the exogenous control of exploratory activity. Their functional reorganization, as a consequence of the infant's participation in intersubjective exchanges, leads to three combined systems: voluntary control of exploratory activity, discrimination of amodal information, and recognition of amodal information. Each of these combined systems results from the reorganization of other systems, natural and/or combined, based on social mediation – as a social system regulating behavior – that gives rise to the exploratory activity that allows the infant to extract temporal amodal information from adult behavior. The adult's infant-directed performance becomes at once a source of temporal multisensory amodal information and an external regulatory activity of the infant's exploratory activity, as we will see below.

Natural System 1. Sensitivity to Amodal Information

The detection or sensitivity to multimodal amodal information is not unique to our species. On the contrary, it is also present in other species some of which are phylogenetically very distant. For example, the male peacock spider (*Maratus volans*) engages in multimodal behaviors when courting a female (Girard et al., 2011). In

quail embryos, exposure to bimodally rendered amodal information determines subsequent discrimination of such information when presented unimodally (Lickliter et al., 2006).

Lewkowicz (2000) outlined some ideas for understanding the detection or sensitivity to temporal amodal information from birth, continuous with intrauterine experience of temporal amodal information. The epigenetic systems theory claims that functions present at birth trace their antecedents to prenatal development. Consequently, and by virtue of the infant's early sensitivity to temporal synchrony, Lewkowicz (1994) considers whether such sensitivity could be responsible for intersensory development during intrauterine life.

We know that the different perceptual systems become functional at different times during intrauterine development (Bremner et al., 2012). The somatosensory and vestibular systems go first, followed by the chemo-sensory systems (oral and nasal) and the auditory and visual systems (Gottlieb, 1971). This sequence restrains the synchronous multisensory experience as it will be experienced outside the uterus until quite late in gestation, when all but the visual systems are active. Beating in mind that much of the research on temporal synchrony in infants has focused on synchrony between visual stimulation and stimulation in other sensory modalities, the fact that the fetus does not have the chance to experience temporal synchrony between the visual and some other stimulation raises a problem: How might prenatal experience contribute, for example, to an early detection of temporal synchrony between auditory and visual information? A possibility might be that prenatal detection of synchrony based on information coming from the (developing) somatosensory, vestibular, and auditory systems works as a forerunner of postnatal detection of intersensory relationships involving visual information (Lewkowicz, 2000).

On the other hand, several empirical studies reveal that toward the last trimester of gestation the human fetus detects (is sensitive to) vibrotactile and acoustic information (Kisilevsky, 1995; Lecanuet et al., 1995) and chemical information (Schaal et al., 1995). Likewise, the fetus produces its own tactile and kinesthetic stimulation as a result of its cyclic motor activity (Bekoff, 1995; Robertson & Bacher, 1995). Human fetuses feel the consequences of their own movement (kinesthesia) simultaneously with the tactile consequences of that movement (e.g., feeling their limbs moving across a body surface such as the uterus). Thus, the duration of information obtained through both systems is equivalent, when considering the synchronous occurrence of the onset of movement and the onset of tactile sensation and the synchrony between the end of movement and the end of tactile sensation. In addition, there is empirical evidence that the fetus detects the rhythmic modeling of somatosensory-tactile-vestibular information in maternal movements (Provasi et al., 2014).

While, as we have seen, intrauterine multisensory experience is possible, its effects are likely to change continuously as perceptual systems mature and the environment changes due to fetal growth and the decrease in the amount of amniotic fluid. However, prenatal multisensory experience provides the basic characteristics for intrauterine developmental sensitivity to amodal information: intensity (the required magnitude for detection to occur), sensitivity to synchrony, and the

temporal dimension of perceptual experiences (duration, rate, and rhythm). Given these conditions, we can postulate a genetic continuity between the fetal multisensory experience and the neonate's early sensitivity to temporal amodal information. Empirical evidence regarding the experience of the human fetus, together with data on the perceptual abilities of neonates and infants less than 2 or 3 months old, leads us to consider sensitivity to amodal temporal information as a natural system.

Natural System 2. The Exogenous Control of Exploration Activity

A change in the infant's exploratory behavior becomes evident around the sixth month of life. This change has been extensively studied and well documented, especially in the development of visual exploratory behavior. Visual attention is displayed over an endogenous or an exogenous stimulus. Endogenous attention, also referred to as top-down or goal-oriented, is presumed to be under the subject's active control, whereas exogenous attention, also referred to as bottom-up or stimulus-controlled, is presumed to be under control of an external stimulus (Chun & Wolfe, 2001). Attention shifts from being exogenous or stimulus-controlled during the first 6 months to being endogenous or actively controlled by the infant by the end of the first year of life (Colombo & Cheatham, 2006; Ruff & Rothbart, 1996).

From the moment of birth, infants tend to orient their perceptual systems toward a particular kind of information. Rivière (1986/2003) called this kind of social orientation *tuning programs*; they consist of a selective preference for parameters of stimulation provided by conspecifics:

> visual stimuli with rounded contours, medium brightness, color, and bulging elements, which are all defining characteristics of the human face [...] as for auditory perception [...] they prefer those sounds with the physical features (e.g., frequency and intensity) that typically define the human voice (Rivière, 1986/2003, p. 113).

In more recent works, newborns' preference for the qualities of biological movement has been identified (Simion et al., 2008). Particularly, with respect to temporal information, it is known that adults resort to a dynamic modeling of duration, rate, and rhythm to attract and sustain the infant's attention toward interaction (M. Martinez 2019). On the other hand, Bahrick and Lickliter (2012, 2014) found that intersensory redundancy guides the infant's attention toward amodal information, as opposed to single modality information, thus facilitating its perception. We shall return to this issue in Sect. 6.7.1.2.

Based on the above data, the exogenous control of the exploratory activity can be considered a natural system. The infant's participation in intersubjective encounters is supported and guided by specific stimuli from individuals, which guide the infant's exploratory activity toward interaction. The stimulus features within the performance developed by the adult during the interaction endogenously control the infant's perceptual activity.

The functional reorganization of the natural systems by means of social mediation leads to the emergence of two combined systems responsible for the three most

prominent abilities of multisensory perception (see Sect. 6.2.2). Detection results from the emergence of the combined system responsible for the voluntary control of the exploratory activity (combined system 1). The ability to discriminate between two patterns of different duration, rate or rhythm is tied to the emergence of the discrimination system (combined system 2). Finally, the ability to establish intersensory relationships based on duration, rate, or rhythm will depend on the emergence of the recognition system (combined system 3).

Combined System 1. Voluntary Control of Exploratory Activity

From both the Gibsonian and the sensorimotor perspectives, perception entails action. In this sense, sensitivity or detection of temporary amodal information alone fails to explain the development of later skills. E. Gibson tentatively explained the emergence of the baby's exploratory activity, out of which perceptual learning takes place based on differentiation. However, as we have already seen, her naturalistic perspective suffers from a number of drawbacks (detailed in the section describing the scope and limits of Gibson's theory). Similarly, one of the major limitations of sensorimotor theory is the lack of an explanation for the genesis of action (also broached in that section). We believe the Vygotskian framework seems appropriate for such an account, as it accounts for the development of higher psychological functions based on the reorganization of elementary functions.

If perception is understood as a sensory-motor contingency linked to objects (O'Reagan and Noë 2001), and of course to people, this contingency depends on the actions of the organism. The sensory-motor theory assumes the existence of patterns of dependence between the sensory stimulus and body movement in a way that mastering the dynamic patterns of stimulus and movement is essential for the perceiving organism (Noë, 2009).

If MPTAI emerges when movement is mastered, its genesis, as a specifically human psychological process, should be explained as the organism's progressive capacity to regulate the movements or actions through which the organism voluntarily creates different contingencies. Vygotsky thought that one criterion to determine the difference between an elementary and a higher function is the shift of the function's locus of control from the environment to the individual (Wertsch, 1985). We suggest that the voluntary control of actions performed with different perceptual systems leading to MPTAI is a combined system resulting from the functional reorganization of two natural systems: detection or sensitivity to amodal information and external control of action. The genesis of MPTAI can be understood as the progressive control of actions generating different sensory-motor contingencies.

Combined System 2. Discrimination of Amodal Information

The discrimination of temporal amodal information concerns the ability to differentiate between two or more distinct patterns of duration, rate, or rhythm. Following the theoretical assumption that perceiving entails the mastery of sensory-motor contingencies (O'Regan & Noë, 2001), discrimination might involve performing two different actions in such a way as to produce two different contingencies.

The exercise of voluntary control of actions is now combined with the natural system responsible for detecting amodal information. Voluntary control, as we will see below, enables the exercise of different actions that generate different sensorymotor contingencies.

Combined System 3. Recognition of Amodal Information

The recognition of amodal temporal information refers to the ability to establish equivalences between the information simultaneously available to two or more perceptual systems based on the same pattern of duration, rate, or rhythm. Discrimination enables the differentiation between an event's dissimilar temporal patterns (duration, rate, rhythm) based on information available to all modalities. On the other hand, recognition allows to identify a relationship between different perceptual systems on the basis of the temporal amodal information that an event provides. Again, if we assume that perceiving involves the awareness of sensorimotor contingencies (O'Regan & Noë, 2001), we may conclude that recognition entails performing two actions with two different perceptual systems in such a way that the same sensorimotor contingency is produced. An action related to one perceptual system allows to discriminate information – a sensorimotor contingency (a given pattern of duration, rate, or rhythm) –, while another action, related to another perceptual system, generates the same sensorimotor contingency.

The synchronous occurrence of information simultaneously available to two perceptual systems is not enough for recognition to occur; it also depends on some particular features of the amodal information, not just any feature of the stimulus (Bahrick and Pick 1994). This idea can be better understood by taking the intersensory preference task as an example. In this task, infants are assumed to recognize intersensory relations when there is evidence that the visual stimulus congruent with the sound stimulus attracts a proportion greater than 0.5 of total gaze time. The assumption states that the infant's visual exploratory activity is at the service of extracting information (auditory and visual) from the environment relative to some kind of perceptual invariant. During the task the infant has to jointly perform an auditory exploratory activity and a visual exploratory activity to discriminate which of two different visual contingencies matches the auditory contingency.

Recognition, as a combined functional system, results from the functional reorganization of two other combined systems, namely discrimination and voluntary exploratory activity. Recognition can be achieved by a self-regulated exploratory

activity guided by the discrimination of a specific quality of the amodal temporal information, be it duration, rate, or rhythm.

6.6.2 *Development of MPTAI Nested in Intersubjective Encounters: Social Mediation*

Up to this point, we have defined each of the natural and combined systems whose functional reorganization makes the ontogenesis of MPTAI explicable. We will now account for the way development takes place based on a reorganization of different systems following social mediation roused by the infant's participation in intersubjective exchanges. Nesting MPTAI development in intersubjective encounters helps to explain the genesis of the voluntary control of exploratory activity related to temporal amodal information. This is a crucial matter if we assume that action is not separated from perception.

First, assuming an interactive perspective avoids the problem of direct realism (Aivar & Travieso, 2009). As noted in Sect. 6.4.4, the information to be extracted by the infant is not simply *there*. Temporal amodal information emerges *within* and *out of* an encounter with others: it results from intersubjective encounters. Since it is offered and highlighted by the adult, temporal amodal information shapes infant-directed adult performance (Español & Shifres, 2015). Our hypothesis states that the development of temporal amodal information as a psychological experience is shaped by intersubjective experience. The presence of temporal amodal information depends on intersubjective contact and intersubjective contact depends on the presence of temporal amodal information (Martínez, 2019). As already mentioned, adult performance is, simultaneously, a source of temporal amodal multisensory information and an external regulatory variable of the infant's exploratory activity. Adult-infant interaction builds on the adult's elaboration and dynamic modeling of temporal amodal information. It should be noted, however, that adult performance is not a preconfigured plan unfolding independently of infant behavior. On the contrary, the adult's repertoire depends on the way the infant participates in the interaction.

Secondly, the interactive perspective is the best approach to explain the genesis of the infant's voluntary action to extract information as a combined functional system. Such voluntary control can be explained by Vygotsky's General Genetic Law of Cultural Development (Vygotsky, 1978). The orientation of perceptual systems is initially controlled by the adult's performance. Adults are inherently attractive stimulus patterns for infants, they dynamically shape amodal information (duration, rate, and rhythm) to attract and sustain the infant's attention towards interaction, since redundantly displayed amodal information attracts the infant's attention and thus facilitates its discrimination (Bahrick & Lickliter, 2012). Subsequently, the infant develops the ability to master actions that generate different sensory-motor contingencies linked to temporal amodal information. The

development of voluntary control of the infant's activity that allows the control of sensory motor contingencies during interaction does not, as assumed by E. Gibson, from a postural control via the organism's maturation; on the contrary, it emerges from the mediation of an external and social system of behavioral regulation.

6.7 Some Remarks on the Ontogenesis of MPTAI

In the previous section, we made a proposal concerning the natural and combined systems whose functional reorganization supports the genesis of MPTAI. In this section, we will first speculate about which are the social mediation systems that enable the functional reorganizations (or combinations) of natural and combined systems, and then we will outline the intersubjective contexts where exploratory activity, discrimination, and perceptual recognition could take place.

6.7.1 Social Mediation Systems Linked to the Functional Reorganization of Psychological Systems

This section addresses the social mediation systems used by adults to externally guide the infant's exploratory activity. These mechanisms are hypothetically responsible for guiding in an *interpsychological* fashion the baby's exploratory activity until its establishment as an intrapsychological function. We should recall at this point the important role assigned to the organism's activity (especially its ontogenesis), especially given the assumption that perception and action are tightly bound together.

According to Stern, as a result of his/her participation in early social play, the infant "has 'got' the temporal patterning of human behavior and the meaning of different changes and variations in tempo and rhythm" (Stern, 2002, p. 21). Social play is a playful activity involving adults and infants whose focus of interest is the other's behavior. Play takes place once other tasks have been completed: there is no breastfeeding, diapering, or bathing, nothing must be done immediately (Stern, 2002). The analysis of the adult's performance during social play allowed to identify some of the structural features closely linked to the regulation of the infant's participation (Dissanayake, 2001; Español & Shifres, 2015; Malloch, 1999/2000; Martínez, 2014; Stern, 2002). We hypothesize that some of these structural features (intersensory redundancy, repetition-variation form, contrast, and affect attunement) function as nonlinguistic semiotic mediation systems, each with a different role, in the development of MPTAI. Adults resort to these semiotic mediation systems on a regular basis during the first year of life seeking to bond with the infant. They do so insofar as they are not brought into play in an intentionally differentiated manner at different infant's ages. In this sense, mediating with different natural or combined

systems at different times during the first year of life would enable functional reorganizations resulting in new combined systems. Our proposal is that, analogous to the way adults use language with children as a semiotic system to control external behavior at later ages, they also use non-linguistic mediation systems, especially during early social play, to regulate infant behavior, particularly those actions performed with the infant's perceptual systems. Consequently, during early social would be the locus of the interpsychological stage in the formation of a process that will later become intrapsychological. We will now describe the semiotic mediation systems we believe to be linked to the development of MPTAI.

Intersensory Redundancy

Intersensory redundancy is the simultaneous and synchronous presence of the same amodal temporal information (duration, rate, or rhythm) available to two or more perceptual systems (Bahrick & Lickliter, 2012). It is very frequently present in early intersubjective exchanges. Our idea is that through intersensory redundancy, the adult regulates infant exploratory activity by selectively orienting it toward temporal amodal information. Although we mentioned earlier that multimodality is present in other species, it has an important value in our species, especially during early social play: the parenting figure can offer temporary amodal information in a unimodal, bimodal, or multimodal way, adjusting it according to his or her expectations regarding the baby's participation.

Bahrick and collaborators (Bahrick & Lickliter, 2012) proposed the intersensory redundancy hypothesis (IRH), a highly suggestive hypothesis that allows us to understand how and under what conditions infants orient their exploratory activity toward amodal as opposed to single modality information. IRH proposes that amodal information stands out above all other properties of an event and is more easily detected when it is multimodally available (present simultaneously for two or more perceptual systems) than when it is unimodally available (present for a single perceptual system). On the other hand, an event's single modality information stands out above other properties and is more easily detected in events that provide unimodal information (available to a single perceptual system).

IRH allows us to understand how adult behavior regulates the infant's exploratory activity as a function of the unimodal, bimodal, or multimodal nature of temporal amodal information. During social play, intersensory redundancy operates as an attractor of the infant's exploratory activity by selectively directing it toward the temporal dimension of the adult's behavior.

The Repetition-Variation Form

Adults often resort to the repetition-variation form as a strategy for reciprocity (Español, 2014b, see Chap. 2). The central idea of the repetition-variation form is that, on the one hand, the adult offers information sufficiently iterative, stable, and

temporally related to his/her own behavior so that the infant can detect it as related to the contingencies that evoke his/her own social responses, while on the other hand, it is also varied enough to sustain the infant's attention and participation (Rivière, 1986/2003). It is worth recalling that according to Piaget, circular reaction implies the repetition of actions aimed at consolidating and executing new schemas. Español (2014b) argues that repetition with variation is the heart of social circular reactions. Thus, the repetition-variation form consolidates those specific actions through which the infant extracts information. The repetition-variation form of adult performance enables the infant to consolidate particular actions that generate particular sensory-motor contingencies.

Contrast

Contrast is a concept borrowed from Garvey's (1977) work on the development of play. In her definition of play, Garvey points out that playful behavior keeps a systematic relationship with non-playful behavior. The contrast between a playful activity and a non-playful activity defines play. Español and Pérez (2015) stress the presence of this experience in the baby's life, which takes place during early social play. According to them:

> This kind of play resembles a dance with a rhythmic and progressive structure, because adults, besides moving the event in time this or that way, express it multimodally: if they lengthen the time, they widen the facial expression, they widen the opening of the eyebrows and eyelids; if they shorten the time, they shorten the distance between their own hands or between the baby's face and theirs. The melodic contours and the kinetic contours of their multimodal performance are coordinated [...] The baby, sensitive to multimodally expressed temporal features [...] perceives and is affected by the delay, the expectation. He/she understands the play and participates. Playing with expectation or suspense, with repetitive cycles and crescendos, generates a full experience of joy and excitement in the baby that could never be achieved by the baby alone, at least not in its cyclical character and its intensity (Español & Pérez, 2015, pp. 15-16).

Contrast is present whenever the adult introduces, by means of variations, a structural difference between the playful and non-playful versions of a given behavior.

Adult-infant interaction takes place in different contexts. They share purposeless situations that are merely playful and therefore inherently unproductive in utilitarian terms –early social play–, or situations where the adult satisfies the child's basic needs like feeding, hygiene, soothing discomfort, and sometimes mixed situations occur (Carretero, 2016). In some sporadic observations, we have detected that when the adult tries to soothe an infant's discomfort, while holding him/her in his/her arms, he/she frequently pats his/her back gently with a steady rate pattern. By contrast, when the adult pats the infant's body during early social play, he/she usually does so by systematically varying the rate pattern. These cases illustrate a structural difference (temporal quality of the patting rate) between a playful and a non-playful behavior.

Affect Attunement

The last social mediation system we will address is affect attunement. Stern (1985) defined affect attunement as a particular type of partial imitation, involving an adult performing an overt behavior different from the infant's where certain features of the original infant behavior are retained and matched. Features to be matched are intensity (absolute or intensity profile), temporal pattern (duration, rhythm, pulse), and spatial pattern. If we specifically focus on the temporal pattern, we can see it depends on temporal amodal information, though Stern refers to rate by the name of "pulse". Adult affect attunement emerges at the child's age of 2 months, and is performed regularly throughout the baby's first year of life (Español et al., 2018, see Chap. 4).

Affect attunement is a behavior whereby the adult offers the infant the same pattern of temporal amodal information that is present in the infant's behavior. It is similar to imitation, affective contagion, and empathy, in that they share the possibility of establishing emotional resonance, while the distinctive feature of affect attunement is that it recasts the emotional experience in some other form of expression, it reformulates a subjective state. The subjective state is considered a referent and the overt behavior is a possible expression of this referent (Español et al., 2018). Stern (2010) understands affect attunement as a behavior through which the adult shares an affective state with the infant that both experience during their interaction, by matching the dynamic forms of his/her actions' vitality with those of the infant. In his original formulation, Stern (1985) understood affect attunement as an occasional behavioral event of the adult that is useful for sharing moment-to-moment micro-changes in intersubjective affective experiences with the infant. With affect attunement, adults offer infants the same temporal amodal information pattern, making it available to a different perceptual system.

Up to this point, we have described the semiotic mediation systems provided by the adult during the interaction. We will now describe a possible genesis of the voluntary control of exploratory activity, the discrimination of temporal amodal information, and the recognition of temporal amodal information as a result of the reorganization of natural and combined systems in the interactive context of early social play.

6.7.2 The Intersubjective Contexts Where Exploratory Activity, Discrimination and Perceptual Recognition Arise

The Genesis of Exploratory Activity: Voluntary Control

The genesis of exploratory activity is the genesis of the voluntary control over actions performed by the organism to extract information from the environment. This point is essential for our explanation, based on Aivar and Travieso's criticism of O'Regan and Noë's (2001) conception, according to which "the model does not

explain how these contingencies generate, the practical knowledge that allows an adequate exercise of our actions" (Aivar & Travieso, 2009, p. 189). In this regard, the genesis of exploratory activity is a core functional system for our model. And as previously mentioned, we believe that the Vygotskian perspective allows us to account for this genesis.

According to Vygotsky (1930/1999), specifically human psychological processes result from the reorganization of elementary processes. The development of the first combined system, the voluntary control of action, results from the reorganization of two natural systems, based on the mediating systems used by the caretakers to regulate infant behavior during interaction. We proposed that infants at birth rely on two natural systems: amodal information sensing and endogenous control of exploratory activity. During intersubjective encounters, adults intuitively use (Papoušek, 1996) a number of resources to attract and sustain the infant's exploratory activity towards interaction. As for the development of exploratory activity, we suggest that mediation systems used by adults are intersensory redundancy, the repetition-variation form, and contrast. From an observer's perspective, we could say that adults use these resources to regulate the infant's mood, and to summon and extend the infant's participation during intersubjective encounters. From the point of view of the reorganization of natural systems (detection of amodal information and exogenous control of exploratory activity), intersensory redundancy, the repetition-variation form, and contrast are at the service of the genesis of the combined system responsible for the voluntary control of exploratory activity.

Exogenous regulation from the adult enables the infant to gradually start to differentiate which perceptual systems are used to produce specific sensory-motor contingencies. From the adult's point of view – during interactions – the infant's behavior is processed through a subjective filter of human interpretation that endows it with coherence and relevance (Newson, 1978), whereby it is understood as a behavior oriented to and motivated by social contact. However, strictly speaking, the infant is just extracting information from the environment. The baby's exogenous control of exploratory activities (by social mediation systems) in the context of early social play, where adults overinterpret his/her behavior, provides an interpsychological context that favors the emergence of endogenous exploratory activity. The progressive development of the endogenous control of exploratory activities leads infants to master the laws underlying the changes (sensory-motor contingencies) corresponding to each perceptual system.

Within the Vygotskian framework, infant behavior is initially organized according to external (social) control through, for example, adult verbal instructions (Díaz et al., 1990). This external regulation method is then internalized by genetically transforming infant behavior (now at an internal level). It is quite implausible that the adult uses verbal instructions to guide the actions performed by the infant with his/her perceptual systems. We postulate instead that the adult uses three nonlinguistic mediation systems to regulate infant exploratory activities: intersensory redundancy, the repetition-variation form, and contrast. Adults mediate between the infant's action (initiation of exploratory activity) and the temporal amodal information (provided by the adult), enabling him/her to master the dynamic patterns of

stimulus and movement to generate specific sensory-motor contingencies. Early social play is the baby's preferred setting to exercise exploratory behaviors that first occur in an interpsychological fashion. In such play, adults elaborate performances where they dynamically model their behavior's duration, rate, and rhythm. Furthermore, these performances carry a historical course of development inherent to the dyad (Carretero, 2016). Intersensory redundancy in these performances attracts the baby's exploratory activity toward temporal amodal information. The repetition-variation form, as the heart of social circular reactions, leads to the establishment, by means of exercise, of specific actions that allow the extraction of information. Exercising a behavior activates and automates collaboration and the composition of functional systems (Rosa et al., 1993). The contrast delivered in adult actions helps the infant to discriminate actions through which different kinds of temporal amodal information are extracted.

Figure 6.1 shows the natural systems and the social mediation mechanisms that account for the emergence of the combined system of voluntary control of exploratory activities. The infant will gradually begin to control voluntarily the actions performed with his/her perceptual systems.

Infants begin to notice which perceptual system generates particular sensory-motor contingencies by voluntarily controlling their actions. O'Regan and Noë (2001) explain how, for example, the sensation of red color arises according to the mastery of the sensory-motor contingencies linked to one's own body, to one's own actions:

> the sensation of red occurs when the brain is tuned to certain very particular things that will happen to the neural influx if we do certain things with respect to the red patch of color. For example, one thing we can do is close our eyes, at which point there should be a very dramatic change in the neural influx. This particular sensorimotor contingency is an indication that we are seeing the red patch of color, and not hearing or tasting it, since eye-closing has no effect in the auditory or tactile modalities (O'Regan & Noë, 2001, p. 83).

Similarly, by mastering the actions performed with his/her perceptual systems, the infant identifies the visual sensation of duration, rate, and rhythm, the auditory

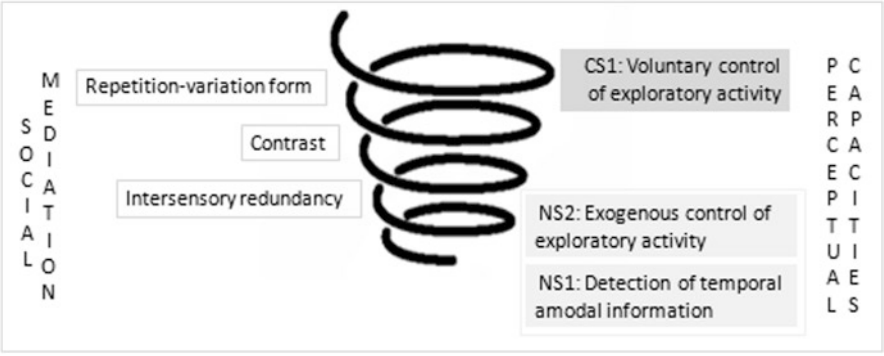


Fig. 6.1 Schematic representation of the genesis of the combined system of voluntary control of exploratory activity

sensation of duration, rate, and rhythm, or the kinesthetic sensation of duration, rate, and rhythm. This is crucial in the case of temporal amodal information, because it can be extracted through any perceptual system. Therefore, it is essential for the development of the organism to master, in an autonomous way, the laws that govern the contingencies by virtue of the actions performed with one's own perceptual systems. Ultimately, this is what allows the combined system of the voluntary control of exploratory activities to develop.

The Genesis of Perceptual Discrimination in the Intersubjective Context

Discrimination is the ability to differentiate between two or more objects, events, or features. In the case of amodal information, it allows to differentiate between two different multimodal patterns of duration, rate, or rhythm that are present in an adult performance. Following the logic proposed by O'Regan and Noë (2001), whereby perception is the mastery of environment-related contingencies, discrimination entails performing an action whose sensory-motor contingency is different from the one produced by the same action previously performed in the face of a change in the temporal pattern of amodal information. The same action generates different contingencies when it is embedded in two different temporal patterns.

E. Gibson proposed that the emphasis on distinctive features and contrast enable perceptual learning. It is essential to describe how and at what rate the infant's (intersubjective) experience takes place, out of which the ability to discriminate develops. Consequently, we must explain how, cradled in intersubjective contact, the functional reorganization of the natural system for the detection of temporal amodal information and the combined system of voluntary control of exploratory activity occur, as a function of social mediation, thus giving rise to the combined system of amodal information discrimination. Social mediation through the repetition-variation form is the best candidate to account for the genesis of the combined system of information discrimination, since it enhances the experience of emphasis of distinctive features. Infant participation in social play involving the repetition-variation form allows us to understand how the experience for instance necessary to develop the ability to discriminate between two different rhythmic patterns of adult performance is produced.

During early social play, adults organize their behavior in the form of repetition-variation. If we take a rhythmically modeled multimodal performance as an example, the behavior will most likely have the following structure: the adult executes a rhythmic pattern A, then repeats the same rhythmic pattern A, and then executes a rhythmic pattern B. The varied repetition exemplifies how, during the intersubjective encounter, the infant experiences emphasis on distinctive features. According to E. Gibson (1969), this experience is necessary for the development of discrimination. The adult offers the A pattern twice, then switches and offers the B pattern, ostensibly different from the first one. During early social play, when the adult conforms to the repetition-variation form, the infant iteratively performs the same actions (with different sensory systems) until a contingency different from those

Table 6.1 The relationship between the adult's behavior and the contingency that generates the infant's action

Adult behavior	Motif A	Repeat motif A	Motif B
	Rhythmic pattern A	Rhythmic pattern A	Rhythmic pattern B
Infant behavior	Action A	Action A	Action B
	Contingency A	Contingency A	Contingency B

Representation of the relationship between the adult's behavior and the contingency that generates the infant's action. The scheme, although exemplified with a rhythmic pattern, works equally well with patterns of duration or rate

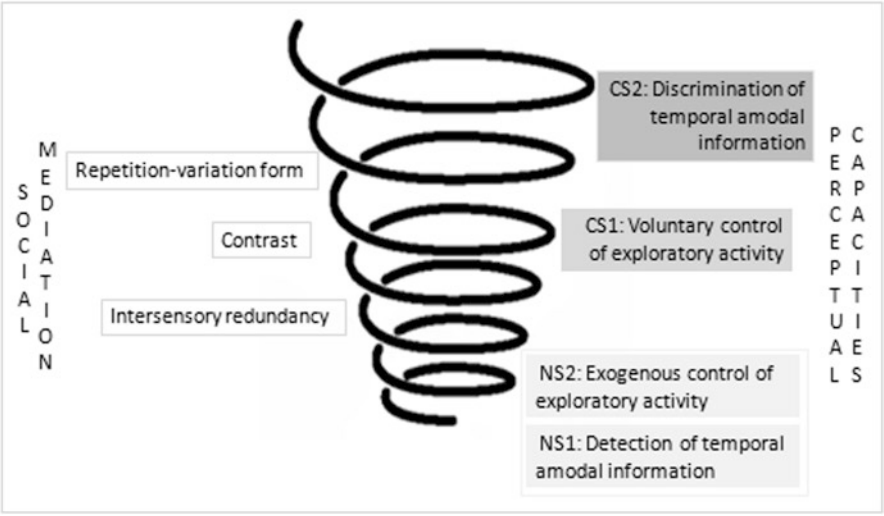


Fig. 6.2 Schematic representation of the genesis of the combined system of discrimination of temporal amodal information

occurring in the previous actions takes place. Discrimination requires the control of voluntary actions so that they can be repeated until the contingency changes. Mastering the law governing changes in contingency (performing the same action) results in the discrimination of different patterns of temporal amodal information. Table 6.1 schematizes the relationship between the repetition-variation form in adult behavior and the actions performed by the infant with their ensuing contingency.

The infant's participation in intersubjective exchanges where the adult organizes his/her behavior with the repetition-variation form ensures the experience of emphasis on distinctive features. The combined system of voluntary control of action is integrated with the natural system of detection of amodal information based on the social mediation of the repetition-variation form, leading to the genesis of a new combined system: discrimination of amodal temporal information. Figure 6.2 shows a diagram identifying the functional systems and the social mediation mechanisms that give rise to discrimination as a functional system.

On the other hand, we must point out that when the infant identifies a change in the temporal pattern that models adult behavior, he/she changes his/her actions (which precisely occurs during his/her participation in the exchange when adult behavior is modeled by the repetition-variation form), extending and reworking the interactive loop shared with the adult during early social play.

The Genesis of Perceptual Recognition in the Intersubjective Context

Following O'Regan and Noë's (2001) formulation, according to which perception is the domain of contingencies linked to the environment, the recognition of intersensory relations involves performing two actions with two different sensory systems until the sensory-motor contingency that each action produces is matched. The recognition of amodal temporal information involves establishing equivalences between the simultaneous information (the same duration, rate, or rhythm pattern) available to two, or more, perceptual systems. Recognition arises from the performance of two different actions leading to the same sensory-motor contingency. Recognition emerges because self-regulated exploratory activity is guided by the possibility of discriminating between different patterns of temporal amodal information.

To account for the emergence of the new combined system requires postulating a social mediation mechanism that guides the infant's exploratory activity (based on the same pattern of temporal amodal information) with respect to the information simultaneously available to two or more perceptual systems. The infant's participation in intersubjective encounters where the adult affectively attunes his behavior may be one of the means of social control that guides exploratory activity according to temporal amodal information (duration, rate, or rhythm). Adults use this information to govern the infant's exploratory activity. The adult provides temporal amodal information in a modality different from that present in the infant's behavior.

Infant participation in intersubjective exchanges where adults use affect attunement as a cultural means to regulate infant behavior leads to the genesis of a new system as a result of the combination of two other combined systems (voluntary control of exploratory activity and discrimination of amodal information). Figure 6.3 depicts the different functional systems and social mediation mechanisms underlying the genesis of recognition. In addition, this figure represents the final model of the ontogenesis of MPTAI.

Infant actions, for example, body movements, generate a sensory-motor contingency with respect to duration, rate, or rhythm, while the auditory or visual exploration of the adult's vocal or motor behavior generates (through different perceptual systems) the same sensory-motor contingency based on temporal amodal information. Recognition, therefore, entails performing actions that generate the same contingency with different perceptual systems.

By way of an example, when the infant recognizes intersensory relations based on temporal amodal information in situations involving affect attunement, he/she alternates his/her exploratory behavior between the information provided by his/her

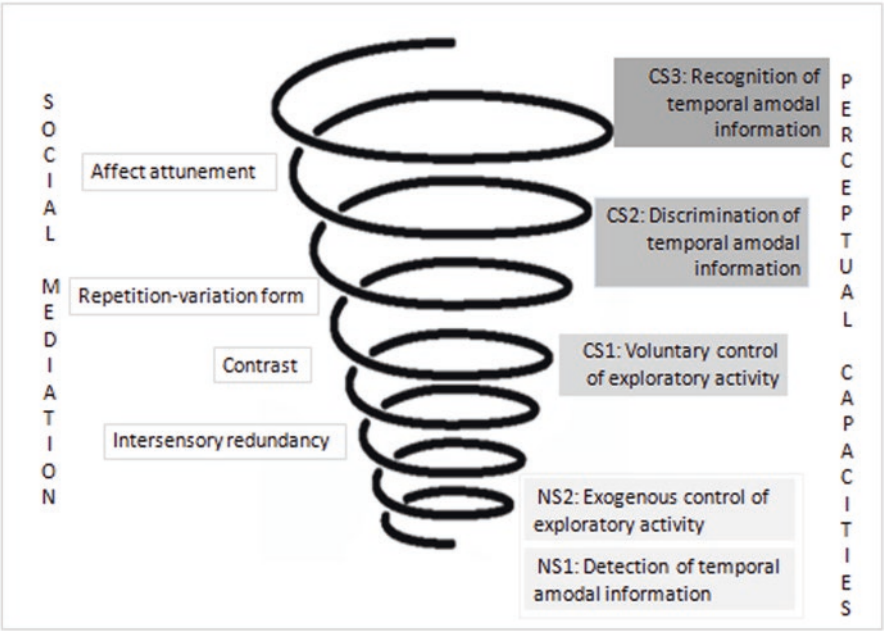


Fig. 6.3 Schematic representation of the genesis of the combined system of recognition of temporal amodal information

own body (for example, the rate at which he/she taps his/her hand on the table of his/her feeding chair) and the one coming from his/her interaction partner (for example, his/her mother's vocalization rate). Another example, taken from laboratory studies, is when the infant exhibits the ability to recognize intersensory relationships: in this case he/she must alternate his/her ocular exploratory activity between two sources of visual information until the action that generates the same sensory-motor contingency as the activity performed with the auditory system becomes stable.

6.8 Finals Remarks

The aim of this chapter was to present a theoretical model to explain the genesis of MPTAI. To do so, we started from two theoretical frameworks (perceptual learning proposed by E. Gibson, and the development of functional psychological systems proposed by Vygotsky) that allowed us to trace a possible genesis of MPTAI, that is, the development of a set of different skills (detection, discrimination, and recognition) linked to different types of temporal amodal information (duration, rate, and rhythm). We have outlined a model of the genesis of MPTAI based on the development of functional systems dependent on different non-linguistic systems of social

mediation. Such cultural systems of external regulation of infant behavior (intersensory redundancy, performative contrast, repetition-variation form, affect attunement) enable the internalization of schemas reflecting the regularities of the actions the infant performs to extract the available temporal amodal information from the environment and its possible intersensory relations when this information is simultaneously accessible to two or more perceptual systems. In this way, the model proposes a possible explanation for the genesis of the endogenous control of exploratory activity. Likewise, considering the development of exploratory activity, it helps to explain the genesis of the discrimination and the recognition of temporal amodal information.

The proposed model provides a way to address two central questions that to the best of our knowledge lack an appropriate explanation. First, as we have already said, it accounts for the genesis of exploratory activity, a central question for the psychological explanation of action coupled with perception (Aivar & Travieso, 2009). Second, the model explains why the discrimination of temporal amodal information precedes the recognition of intersensory relations based on temporal amodal information. For example, discrimination of bimodal rhythmic patterns occurs around the age of 7 months, while recognition of rhythm-based intersensory relationships emerges around the 10th month (Martínez et al., 2018). Bahrick and Pickens (1994) argue that discrimination precedes recognition because the habituation/dishabituation task (used to study discrimination) requires less attentional capacity and less sophisticated cognitive skills than the cross-modal preference task (used to study recognition). This explanation, however, does not account for the genesis of the attentional capacities or cognitive abilities that underlie discrimination or recognition. In contrast, our model does explain why discrimination precedes recognition. This is because the genesis of recognition depends on the prior development of the combined functional system of discrimination of amodal information.

Although the model, as mentioned in the previous paragraph, explains the genesis of MPTAI by addressing some issues that were previously not properly tackled, it has certain limitations. Some of these may be overcome as more empirical evidence regarding the development of particular multisensory perception skills becomes available. For example, we lack sufficient empirical information on the development the detection and discrimination of temporal amodal information based on duration and rate. Other limitations may be circumvented as the role of the proposed mediating systems is empirically investigated in more detail. For example, the infant's exploratory behavior in relation to the temporal quality of semiotically mediated amodal information through contrast (see Sect. 6.7.1.4). Consequently, the first limitation that should be addressed in subsequent formulations of the model is to adjust the model to fit more accurately the particular developmental sequence of detection, discrimination, and recognition of temporal amodal information. While the model fits the existing data on the above-mentioned sequence with respect to rhythm, it does not conform to duration and rate. This is partly due to the fact that, as mentioned above, we still need further empirical information to establish the sequence.

By linking perceptual and intersubjective development, our model raises new pertinent research questions related to adult-infant performance and early intersubjective development beyond the realm of multisensory perception. These could provide feedback for a better understanding of MPTAI and early intersubjective development. For instance, it would be relevant to investigate how adults regulate infant activity during intersubjective encounters (through different non-linguistic mediation systems included in the model), and the dynamic modeling of their performance in terms of duration, rate, and rhythm.

References

Adolph, K. E., & Kretch, K. S. (2015). Gibson's theory of perceptual learning. *International encyclopedia of the social and behavioral sciences, 10*, 127–134. https://doi.org/10.1016/B978-0-08-097086-8.23096-1

Aivar, M. P., & Fernández, T. R. (2006). La dualidad percepción-acción: la necesidad de recuperar la crítica de Dewey a la teoría del arco reflejo [The perception-action duality: The need to recover Dewey's critique of the reflex arc theory]. *Revista de Historia de la Psicología, 27*(2–3), 269–277.

Aivar, M.P. & Travieso, D. (2009). Percepción, acción y sujeto [Perception, action and subject]. In J.C. Loredo Narciandi, T. Sanchez-Criado & D. Lopez Gómez (eds.), *¿Donde reside la accion? Agencia, Constructivismo y Psicologia* [Where does action reside? Agency, Constructivism and Psychology] (pp. 149–198). Editum y UNED.

Aivar, M. P., Fernandez, T. R., & Sánchez, J. C. (2002). Sensación y Movimiento: Un analisis histórico en torno a la obra de J. J. Gibson. [Sensation and Movement: A Historical Analysis of J. J. Gibson's Work]. *Revista de Historia de la Psicología, 23*(3–4), 381–394.

Allen, T. W., Walker, K., Symonds, L., & Marcell, M. (1977). Intrasensory and intersensory perception of temporal sequences during infancy. *Developmental Psychology, 13*(3), 225–229. https://doi.org/10.1037/0012-1649.13.3.225

Bahrick, L. E. (1992). Infants' perceptual differentiation of amodal and modality-specific audio-visual relations. *Journal of Experimental Child Psychology, 53*, 180–199. https://doi.org/10.1016/0022-0965(92)90048-B

Bahrick, L. E., (1994). The development of infants' sensitivity to arbitrary intermodal relations. Ecological Psychology, 6(2), 111–123. https://doi.org/10.1207/s15326969eco0602_2

Bahrick, L. E., (2001). Increasing specificity in perceptual development: Infants' Detection of nested levels of multimodal stimulation. Journal of Experimental Child Psychology, 79(3), 253–270. https://doi.org/10.1006/jecp.2000.2588

Bahrick, L. E. (2004). The development of perception in a multimodal environment. In G. Bremner & A. Slater (Eds.), *Theories of infant development* (pp. 90–120). Blackwell Publishing. https://doi.org/10.1002/9780470752180.ch4

Bahrick, L. E. (2009). Perceptual development: Amodal perception. In B. Goldstein (Ed.), *Encyclopedia of perception* (Vol. 1, pp. 44–46). Sage Publishers.

Bahrick, L. E. (2010). Intermodal perception and selective attention to intersensory redundancy: Implications for typical social development and autism. In G. Bremner & T. D. Wachs (Eds.), *The Wiley-Blackwell handbook of infant development* (2nd ed., pp. 120–166). Blackwell Publishing. https://doi.org/10.1002/9781444327564.ch4

Bahrick, L. E., & Lickliter, R. (2009). Perceptual development: Intermodal perception. In B. Goldstein (Ed.), Encyclopedia of Perception, Vol. 2 (pp. 753–756). Newbury Park, CA: Sage Publishers.

Bahrick, L. E., & Hollich, G. (2017). Intermodal Perception. In *Reference module in neuroscience and biobehavioral psychology* (pp. 1–15). Elsevier.

Bahrick, L. E., & Lickliter, R. (2004). Infants' perception of rhythm and tempo in unimodal and multimodal stimulation: A developmental test of the intersensory redundancy hypothesis. *Cognitive, Affective, & Behavioral Neuroscience, 4*, 137–147. https://doi.org/10.3758/CABN.4.2.137

Bahrick, L. E., & Lickliter, R. (2012). The role of intersensory redundancy in early perceptual, cognitive, and social development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 183–205). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0008

Bahrick, L. E., & Lickliter, R. (2014). Learning to attend selectively: The dual role of intersensory redundancy. *Current Directions in Psychological Science, 23*, 414–420. https://doi.org/10.1177/0963721414549187

Bahrick, L. E. & Pickens, J. N. (1994). Amodal relations: The basis for intermodal perception and learning. En D. Lewkowicz & R. Lickliter (Eds.), The development of intersensory perception: Comparative perspectives (pp. 205–233). : Lawrence Erlbaum Associates.

Bekoff, A. (1995). Development of motor behavior in chick embryos. In J. P. Lecanuet, W. P. Filer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 191–204). Erlbaum.

Blanco, F. (2002). *El cultivo de la Mente: un ensayo histórico-crítico sobre la cultura psicológica* [The cultivation of the Mind: a historical-critical essay on psychological cultura]. Antonio Machado Libros.

Bordoni, M., Español, S., & De Grande, P. (2016). La incidencia del entonamiento afectivo y la imitación en el involucramiento visual-social temprano [The Incidence of Affect Attunement and Imitation in Early Social-Visual Engagement]. *Avances en Psicología Latinoamericana, 34*(3), 487–503. https://doi.org/10.12804/apl34.3.2016.04

Bower, T. G. R. (1974). *Development in infancy* [El desarrollo del niño pequeño (M. Carretero and V. García-Hoz Rosales, trads.). Madrid: Editorial Debate, 1979]. Freeman & Co.

Bremner, A. J., Lewkowicz, D. J., & Spence, C. (2012). The multisensory approach to development. In A. Bremner, D. J. Lewkowicz, & C. Spence (Eds.), *Multisensory development* (pp. 1–26). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199586059.003.0001

Bremner, J. G., Dunn, K. J., & Tham, D. S. Y. (2017). Perception, knowledge, and action. In A. Slater & J. G. Bremner (Eds.), *An introduction to developmental psychology* (3rd ed., pp. 143–182). Wiley.

Brich, H. G., & Lefford, A. (1963). Intersensory development in children. *Monographs of the Society for Research in Child Development., 28*(5), 1–47. https://doi.org/10.2307/1165681

Carretero, S. (2016). *El contacto corporal en interacciones de musicalidad comunicativa temprana* [Body contact in early communicative musicality interactions]. Unpublished doctoral dissertation. Universidad Nacional de Córdoba.

Chun, M. M., & Wolfe, J. M. (2001). Visual attention. In B. Goldstein (Ed.), *Blackwell handbook of sensation and perception* (pp. 272–310). Blackwell Publishing Ltd.

Colombo, J., & Cheatham, C. L. (2006). The emergence and basis of endogenous attention in infancy and early childhood. In R. V. Kail (Ed.), *Advances in child development and behavior* (Vol. 34, pp. 283–322). Elsevier Academic Press.

de la Jacquier, M. P., Martínez, G., Pereira Ghiena, A. & Silva, V. (2013). La organización rítmica de la música [The rhythmic organization of music]. In F. Shifres & M. I. Biurcet (eds.) *Escuchar y pensar la música: Bases teóricas y metodológicas* [Listening and thinking about music: Theoretical and methodological basis] (pp. 172–204). EDULP.

Dember, W. N., & Warm, J. S. (1979). *Psychology of perception* (2nd ed.) [Psicología de la percepción (C. González, trad.). Madrid: Alianza Editorial, 1990]. Holt, Rinehart, & Winston.

Díaz, R., Neal, C., & Amaya-Williams, M. (1990). The social origins of self-regulation. In L. Moll (Ed.), *Vygotsky and education: Instructional implications and applications of Sociohistorical psychology* (pp. 127–154) [Vygotsky y la educación. Connotaciones y aplicaciones de la

psicología sociohistórica en la educación (E. Sinnot, trad, B. Braslavsky, sup.). Buenos Aires: Aique, 1993, pp. 153–185]. Cambridge University Press. https://doi.org/10.1017/CBO9781139173674.007

Dissanayake, E. (2001). Becoming homo Aestheticus: Sources of aesthetic imagination in mother-infant interactions. *SubStance, 30*(1/2), 85–103. https://doi.org/10.1353/sub.2001.0005

Español, S. (2003). De la emoción al espíritu metafórico. Semiosis e intersubjetividad en el desarrollo humano [From emotion to metaphorical spirit: Semiosis and intersubjectivity in human development]. *Studies in Psychology, 24*(3), 277–311. https://doi.org/10.1174/021093903770411193

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana* [How to do things without words. Gesture and fiction in early infancy]. Antonio Machado Libros.

Español, S. (2010a). Interazione precoce. Una prospettiva vygotskijana a partire dagli schemi di Piaget, 25 anni dopo [Early interaction. A Vygotskian perspective from Piaget's schemes, 25 years later]. *Metis, 17*(1), 67–91.

Español, S. (2010b). El desarrollo como estrategia adaptativa: características exclusivas de la infancia humana [Development as an adaptive strategy: characteristics unique to human infancy]. *Revista De Psicología, 11*, 47–58.

Español, S. (2014a). *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality]. Paidós.

Español, S. (2014b). La forma repetición-variación: una estrategia para la reciprocidad [The repetition-variation form: a strategy for reciprocity]. In S. Español (ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality] (pp.157–192). Paidós.

Español, S., & Pérez, D. (2015). Los juegos, los contrastes y las artes [Kinds of play, contrasts, and arts]. *Boletín de estética, 33*, 5–40.

Español, S., & Shifres, F. (2015). The artistic infant directed performance: A microanalysis of the adult's movements and sounds. *Integrative Psychological and Behavioral Science, 49*(3), 371–397. https://doi.org/10.1007/s12124-015-9308-4

Español, S., Bordoni, M., Carretero Pérez, S., Martínez, M., & Camarasa, R. (2018). Imitación y entonamiento afectivo en el juego social temprano [Imitation and affect attunement in early social play]. *Interdisciplinaria: Revista de Psicología y Afines., 35*(2), 291–305.

García, L. (2016). Hacia el canon: Vygotski en España y Argentina (1978–1991) [Towards the Canon: Vygotski in Spain and Argentina (1978–1991)]. In A. Yasnitsky, R. van der Veer, E. Aguilar & L. García (eds.). *Vygotski revisitado. Una historia crítica de su contexto y legado* [Vygotski revisited. A critical history of his context and legacy] (pp.337–359), Miño y Dávila Editores.

García-Albea, J. (1993). *Mente y conducta. Ensayos de psicología cognitiva* [Mind and behavior. Essays in cognitive psychology]. Editorial Trotta.

Garvey, C. (1977). *Play* [El juego infantil (A. Guerra Miralles, trad.) Madrid: Morata. 1985, 4ta. Edición]. Open Books Publishing, Ltd.

Gibson, J. (1966). *The senses considered as perceptual systems*. Greenwood Press.

Gibson, E. (1969). Principles of perceptual learning and development. *New York*. Appleton–Century–Crofts.

Gibson, E. (1970). The development of perception as an adaptive process. *American Scientist, 58*, 98–107.

Gibson, J. (1979). *The ecological approach to visual perception*. Houghton Mifflin.

Gibson, E. (1988). Exploratory behavior in the development of perceiving, acting, and the acquiring of knowledge. *Annual Review of Psychology, 39*, 1–41. https://doi.org/10.1146/annurev.ps.39.020188.000245

Gibson, E. (1997). An ecological psychologist's prolegomena for perceptual development: A functional approach. In C. Dent-Read, P. Zukow-Goldring, & P. (Eds.), *Evolving explanations of*

development: Ecological approaches to organism-environment systems (pp. 23–45). American Psychological Association.
Gibson, J., & Gibson, E. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62*(1), 32–41. https://doi.org/10.1037/h0048826
Gibson, E., & Pick, A. (2000). *An ecological approach to perceptual learning and development*. Oxford University Press.
Girard, M. B., Kasumovic, M. M., & Elias, D. O. (2011). Multi-modal courtship in the peacock spider, Maratus volans (OP-Cambridge, 1874). *PLoS One, 6*(9), c25390. https://doi.org/10.1371/journal.pone.0025390
Goldstone, R. L. (1998). Perceptual learning. *Annual Review of Psychology, 49*(1), 585–612. https://doi.org/10.1146/annurev.psych.49.1.585
Gottlieb, G. (1971). Ontogenesis of sensory function in birds and mammals. In E. Tobach, L. R. Aronson, & E. Shaw (Eds.), *The biopsychology of development* (pp. 67–128). Academic Press.
Horowitz, F., Paden, L., Bahana, K., & Self, P. (1972). An infant–control procedure for studying infant visual fixations. *Developmental Psychology, 7*, 90. https://doi.org/10.1037/h0032855
Kisilevsky, B. S. (1995). The influence of stimulus and subject variables on human fetal responses to sound and vibration. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 263–278). Erlbaum.
Koester, L. S., Papoušek, H., & Papoušek, M. (1989). Patterns of rhythmic stimulation by mothers with three-month-olds: A cross-modal comparison. *International Journal of Behavioral Development, 12*(2), 143–154. https://doi.org/10.1177/016502548901200201
Lecanuet, J. P., Granier-Deferre, C., & Busnel, M. C. (1995). Human fetal auditory perception. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 239–262). Erlbaum.
Lewkowicz, D. J. (1986). Developmental changes in infants' bisensory response to synchronous durations. *Infant Behavior and Development, 9*(3), 335–353. https://doi.org/10.1016/0163-6383(86)90008-1
Lewkowicz, D. J. (1994). Development of intersensory perception in human infants. In D. J. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception. Comparative perspectives* (pp. 165–203). Lawrence Erlbaum Associates.
Lewkowicz, D. J. (2000). The development of intersensory temporal perception: An epigenetic systems/limitations view. *Psychological Bulletin, 126*(2), 281–308. https://doi.org/10.1037/0033-2909.126.2.281
Lewkowicz, D. J. (2003). Learning and discrimination of audiovisual events in human infants: The hierarchical relation between intersensory temporal synchrony and rhythmic pattern cues. *Developmental Psychology, 39*(5), 795. https://doi.org/10.1037/0012-1649.39.5.795
Lewkowicz, D. J. (2012). Development of multisensory temporal perception. In M. M. Murray & M. T. Wallace (Eds.), *The neural bases of multisensory processes* (pp. 325–344). CRC Press.
Lewkowicz, D. J., & Marcovitch, S. (2006). Perception of audiovisual rhythm and its invariance in 4-to 10-month-old infants. *Developmental Psychobiology, 48*(4), 288–300. https://doi.org/10.1002/dev.20140
Lickliter, R., Bahrick, L. E., & Markham, R. G. (2006). Intersensory redundancy educates selective attention in bobwhite quail embryos. *Developmental Science, 9*, 604–615. https://doi.org/10.1111/j.1467-7687.2006.00539.x
Luria, A. R. (1973). *The working brain. An introduction to neuropsychology* [El cerebro en acción (M. Torres, trad.) Madrid: Martínez Roca, 1988]. Penguin Books.
Malloch, S. N. (1999/2000). Mothers and infants and communicative musicality. *Musicae Scientiae, 3*(1 suppl), 29–57. https://doi.org/10.1177/10298649000030S104
Mareschal, D., Westermann, G., & Althaus, N. (2012). In search of the mechanism of multisensory development. In A. Bremner, D. Lewkowicz, & C. Spence (Eds.), *Multisesnory development* (pp. 342–359). Oxford University Press.

Martí, E. (1996). The mechanisms of internalisation/externalisation of knowledge in Piaget's and Vygotsky's theories. In A. Tryphon & J. Vonèche (eds.), *Piaget-Vygotsky. The social genesis of thouht* [Piaget-Vygotsky. La génesis social del pensamiento (J. Piatigorsky, Trad.). Buenos Aires: Argentina, 2000, pp. 81–113]. Taylor & Francis Psychologu Press.

Martínez, I. (2014). La base corporeizada del significado musical [The embodied basis of musical meaning]. In S. Español (ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* [Music and developmental psychology. An interdisciplinary exploration of human musicality]. (pp.71–110). Paidós.

Martínez, M. (2016). Desarrollo de la percepción intersensorial. Aspectos metodológicos, empíricos y conceptuales [Development of intersensory perception. Methodological, empirical and conceptual aspects.]. In E. Huaire Inacio, Á. Elgier & G. Clerici (eds.), *Pensar la niñez: Psicología del desarrollo desde una perspectiva americana* [Thinking childhood: Developmental psychology from an American perspective] (pp. 39–59). Editora y Librería Jurídica Grijley EIRL.

Martínez, M. (2019). Desarrollo intersubjetivo y perceptivo: el ritmo como ejemplo de su enlace durante el primer año de vida [Intersubjective and perceptual development: Rhythm as an example of its link during the first year of life].In C. Scotto, F. Rodríguez & I. Audisio (eds.), *Los signos del cuerpo. Enfoques multimodales de la mente y el cuerpo* [The signs of the body. Multimodal approaches to mind and body]. (pp. 273–325). TESEO-UAI.

Martínez, M., Español, S., & Igoa, J. M. (2018). Reconocimiento de relaciones intersensoriales basadas en el ritmo a los 4, 7 y 10 meses [Recognition of rhythm–based intersensory relationships at 4, 7, and 10 months]. *Epistemus, Revista de Estudios en Música, Cognición y. Cultura, 6*(2), 10–32. https://doi.org/10.21932/epistemus.6.6233.2

Martínez, M., Español, S., & Igoa, J. M. (2022). Recognition of intersensory relationships based on duration in infants four, seven and 10 months old. *Journal for the Study of Education and Development, 45*(1), 81–115. https://doi.org/10.1080/02103702.2021.1925457

Newson, J. (1978). Dialogue and development. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 31–43). Academic Press.

Noë, A. (2009). *Out of our heads. Why you ar not your brain, and other lessons from the biology of consciousness* [Fuera de la cabeza. Por qué no somos el cerebro, y otras lecciones de la biología de la conciencia (N. d'Amonville, trad. A. Conde, rev.) Barcelona: Kairós, 2010]. Hill & Wang.

O'Regan, J. K., & Noë, A. (2001). What it is like to see: A sensorimotor theory of perceptual experience. *Synthese, 129*, 79–103. https://doi.org/10.1023/A:1012699224677

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 88–112). Oxford University Press.

Pereira Ghiena, A., de la Jacquier, M. P., Silva, V. & de la Saint Pierre, M. P. (2013). La Estructura Métrica [Metrical structure]. In F. Shifres & M. I. Burcet (eds.) *Escuchar y pensar la música: Bases teóricas y metodológicas* [Listening and thinking about music: Theoretical and methodological basis] (pp. 153–170). EDULP.

Piaget, J. (1977). *La construction du réel chez l'enfant* [La construcción de lo real en el niño (R. Santamaría, trad.) Buenos Aires: Grijalbo, 1988]. Delachaux & Niestlé S.A.

Pickens, J. N., & Bahrick, L. E. (1995). Infants' discrimination of events on the basis of rhythm and tempo. *British Journal of Developmental Psychology, l3*, 223–236. https://doi.org/10.1111/j.2044-835X.1995.tb00676.x

Provasi, J., Anderson, D. I., & Barbu-Roth, M. (2014). Rhythm perception, production, and synchronization during the perinatal period. *Frontiers in Psychology, 5*, 1048. https://doi.org/10.3389/fpsyg.2014.01048

Read, C., & Szokolszky, A. (2018). An emerging developmental ecological psychology: Future directions and potentials. *Ecological Psychology, 30*(2), 174–194. https://doi.org/10.1080/10407413.2018.1439141

Real Academia Española (2019). Duración. En *Diccionario de la lengua* española (23.3 ed.). Retrieved from: https://dle.rae.es/?w=duración

Rivère, Á. (1999/2003). Desarrollo y Educación: El papel de la educación en el "diseño" del desarrollo humano [Development and education: The role of education in the "design" of human development]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comps.) *Ángel Rivière. Obras escogidas*, Vol. III [Ángel Rivière. Collected Works, Vol, III] (pp. 203–242). Panamericana.

Rivière, Á. (1984/2003). Acción e interacción en el origen del símbolo [Action and interaction at the origin of the symbol]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comps.) *Ángel Rivière. Obras escogidas*, *Vol. II* [Ángel Rivière. Collected Works, Vol, II] (pp.77–108). Panamericana.

Rivière, A. (1986/2003). Interacción precoz. Una perspectiva vygotskiana a partir de los esquemas de Piaget [Early interaction. A Vygotskian perspective from Piaget's schemas]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comp.) *Ángel Rivière. Obras Escogidas, Vol II* [Ángel Rivière. Collected Works, Vol, II] (pp. 109–142). Panamericana.

Rivière, A., & Español, S. (2003). La suspensión como mecanismo de creación semiótica [in English: Suspension as a semiotic creation mechanism]. *Studies in Psychology, 24*(3), 261–275. https://doi.org/10.1174/021093903770411184

Rivière, Á. & Sotillo, M. (1999/2003). Comunicación, suspensión y semiosis humana: Los orígenes de de la práctica y de la comprensión interpersonales [Communication, suspension and human semiosis: The origins of interpersonal practice and understanding]. In M. Belinchón, A. Rosa, M. Sotillo & I. Marichalar (comps.) *Ángel Rivière Obras escogidas, Vol III* [Ángel Rivière. Collected Works, Vol, III] (pp.181–201). Panamericana.

Robertson, S. S., & Bacher, L. F. (1995). Oscillation and chaos in fetal motor activity. In J. P. Lecanuet, W. P. Fifer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 169–189). Erlbaum.

Rochat, P. (2001). *The infant's world* [El mundo del bebé (R. Filella, trad.). Madrid: Morata, 2004]. Harvard University Press.

Rodríguez, F. G., & Español, S. (2016). Composiciones bimodales de gesto y vocalización en el inicio de la comunicación verbal [Bimodal compositions of gesture and vocalization at the beginning of verbal]. *Journal for the Study of Education and Development, 39*(4), 661–693. https://doi.org/10.1080/02103702.2016.1215086

Rosa, A., Huertas, J. A. & Blanco, F. (1993). Psicología de la ceguera y psicología general [Psychology of blindness and general psychology]. In A. Rosa & E. Ochaita (Eds.), *Psicología de la ceguera* [Psychology of blindness] (pp. 319–361). Alianza Editorial.

Ruff, H. A., & Rothbart, M. K. (1996). *Attention in early development. Themes and variations*. Oxford University Press.

Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In D. Kuhn & R. Siegler (Eds.), *Handbook of child psychology* (Vol. 2). Wiley.

Schaal, B., Orgeur, P., & Rognon, C. (1995). Odor sensing in the human fetus: Anatomical, functional, and chemoecological bases. In J. P. Lecanuet, W. P. Filer, N. A. Krasnegor, & W. P. Smotherman (Eds.), *Fetal development: A psychobiological perspective* (pp. 205–237). Erlbaum.

Simion, F., Regolin, L., & Bulf, H. (2008). A predisposition for biological motion in the newborn baby. *Proceedings of the National Academy of Sciences, 105*(2), 809–813. https://doi.org/10.1073/pnas.0707021105

Slater, A., Riddell, P., Quinn, P. C., Pascalis, O., Lee, K., & Kelly, D. J. (2010). Visual perception. In J. G. Bremner & T. D. Wachs (Eds.), *The Wiley-Blackwell handbook of infant development* (pp. 40–80). Blackwell Publishing Ltd.

Spelke, E. (1976). Infants' intermodal perception of events. *Cognitive Psychology, 8*(4), 553–560. https://doi.org/10.1016/0010-0285(76)90018-9

Stern, D. N. (1985). *The interpersonal world of the infant. A view from Psychoanalysis and development psychology* [El mundo interpersonal del infante. Una perspectiva desde el psicoanálisis

y la psicología evolutiva (J. Piatigorsky, trad.). Buenos Aires: Paidós, 1991] Basic Books, Inc. Publishers.

Stern, D. N. (2002). *The first relationship: Infant and mother*. Harvard University Press.

Stern, D. N. (2010). *Forms of vitality: Exploring dynamic experience in psychology, the arts, psychotherapy, and development*. Oxford University Press.

Szokolszky, A., & Read, C. (2018). Developmental ecological psychology and a coalition of ecological–relational developmental approaches. *Ecological Psychology, 30*(1), 6–38. https://doi.org/10.1080/10407413.2018.1410409

Tees, R. (1994). Early stimulation history, the cortex, and intersensory functioning in infrahumans: Space and time. In D. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory percpeption. Comparatives* (pp. 107–131). Lawrence Erlbaum Associates.

Travieso, D. (2000). *Alteraciones funcionales del tacto en la conjunción entre Ceguera y Diabetes* Mellitus [Functional alterations of touch in the conjunction between blindness and diabetes mellitus]. Unpublished doctoral dissertation. Universidad Autónoma de Madrid.

Travieso, D. & Blanco, F. (1999). Alteraciones perceptivas [Perceptual impairments]. In E. Munar, J. Rosselló & A. Sánchez-Cabaco (Coords.), *Atención y percepción* [Attention and perception] (pp.629–657). : Alianza Editorial.

Travieso, D., & Blanco, F. (2000). Las teorías ecológicas y soviética de la percepción: un estudio crítico sobre la relación entre sus teorías sobre el tacto [Ecological and soviet theories of perception: A critical study on the relationship between their theories of touch]. *Revista de Historia de la Psicología, 2-3*(21), 401–412.

Valsiner, J. (1998). The development of the concept of development: Historical and epistemological perspectives. In W. Damon & R. Lerner (Comp.), Handbook of child Pschicology (pp.189–232). : Willey.

Van der Veer, R., & Van Ijzendoorn, M. H. (1985). Vygotsky's theory of the higher psychological processes: Some criticisms. *Human Development, 28*(1), 1–9. https://doi.org/10.1159/000272931

Varela, F., Thompson, E., & Rosch, E. (1991). *The embodied mind. Cognitive science and human experience* [De cuerpo presente. Las ciencias cognitivas y la experiencia humana (C. Gardini, trad.). Barcelona: Gedisa, 2005). MIT Press.

Vygotsky, L. S. (1930a). On psychological systems. In R. W. Rieber & J. Wollock (Eds.), *The collected works of L. S. Vygotsky, Volume 3. Problems of the theory and history of psychology* (pp. 91–107). Plenum Press. https://doi.org/10.1007/978-1-4615-5893-4

Vygotsky, L. S. (1930b). Tool and sing in the development of the child. In R. W. Rieber (Ed.), *The collected works of L. S. Vygotsky, Volume 6. Scientific legacy* (pp. 3–68). Kluwer Academic Publishers. https://doi.org/10.1007/978-1-4615-4833-1

Vygotsky, L. S. (1932a). Peception and its development in childhood. In R. W. Rieber & A. S. Carton (Eds.), *The collected works of L. S. Vygotsky, Volume 1. Problems of general psychology* (pp. 289–300). Plenum Press. https://doi.org/10.1007/978-1-4613-1655-8

Vygotsky, L. S. (1932b). Infancy. In R. W. Rieber (Ed.), *The collected works of L. S. Vygotsky, Volume. 5. Child psychology* (pp. 207–241). Plenum Press. https://doi.org/10.1007/978-1-4615-5401-1_7

Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological functions* [El desarrollo de los procesos psicológicos superiores (S. Furió, trad.). Barcelona: Crítica, 1979]. Harvard University Press.

Walk, R. D., & Gibson, E. J. (1961). A comparative and analytical study of visual depth perception. *Psychological Monographs: General and Applied, 75*(15), 1–44. https://doi.org/10.1037/h0093827

Walker-Andrews, A. (1994). Taxonomy for intermodal relations. In D. Lewkowicz & R. Lickliter (Eds.), *The development of intersensory perception: Comparative perspectives* (pp. 39–56). LEA.

Walker-Andrews, A. S. (1997). Infants' perception of expressive behaviors: Differentiation of multimodal information. *Psychological Bulletin, 121*(3), 437. https://doi.org/10.1037/0033-2909.121.3.437

Werner, H. (1948). *Comparative psychology of mental development* [Psicología comparada del desarrollo mental (E. Rodríguez Daviro, trad., V. Willams, sup.) Buenos Aires: Paidós, 1965]. International Universities Press
Wertsch, J. (1985). *Vygotsky and the social formation of mind*. Harvard University Press.

Mauricio Martínez has a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). He participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning at the Faculty of Latin America Social Sciences (FLACSO). He is Professor of Developmental Psychology at Universidad Abierta Interamericana (UAI) and Faculty member of the MsC. in Cognitive Psychology (FLACSO). He also teaches in the post-graduate virtual course The baby's life: psychological development, scaffolding and nurturing environmen" at FLACSO. His research focuses on early perceptual and intersubjective development, childhood play and early intervention in autism spectrum disorders. He is author of Intervención psicoeducativa para niños con trastorno del espectro autista [Psychoeducational intervention for children with autism spectrum disorders] (2015). His publications are available at: https://www.researchgate.net/profile/Mauricio-Martinez-2

Chapter 7
The Transition from Early Bimodal Gesture-Word Combinations to Grammatical Speech

Fernando G. Rodríguez and Silvia Español

Abstract Among the first baby communications, we find the bimodal collaborations of gesture and word. In this chapter, we present some considerations on the types of baby gestures and propose a comprehensive observation category system of gestures and vocalizations, either as single signs or in bimodal compositions, based on a relational-semantic pattern. We believe that such a system could bridge the gap between the study of pregrammatical and grammatical communicative expressions. In fact, a multimodal combinatorial role semantics would be better able than other systems to reliably track how simple meanings begin to interweave with one another to express more complex contents within a single message. In this way, the transition from the original multimodal expression to the grammatical form, with its tendency toward unimodality, could be approached in a seamless and integrated manner. In addition, we highlight the pending challenge of expanding the margins of bimodal combinations. In this direction, we point out the need to enrich these categories with the creation of an ecological code of the baby's integral communication resources, which includes both those signs effectively used in communicative acts and the behavioral variables that accompany these signs and guide the interlocutor with keys for their interpretation.

Keywords Gesture · Vocalization · Bimodal combination · Semantic roles · Ecological multimodality · Social cognition

F. G. Rodríguez (✉)
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI)), Buenos Aires, Argentina

S. Español
Latin American Institute of Social Research (IICSAL), Latin American Faculty of Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_7

7.1 Introduction

From the beginning of life, the human world is multimodal: infants receive information through their different perceptual channels and express and communicate through different modalities. Communication varies throughout development. There is a particular form of communication that takes place when the child first combines gesture and word. The approach to bimodal gesture-word combinations in early infancy has refreshed current ideas concerning the ontogenetic development of semiosis. In this chapter, we intend to broaden its study by incorporating a semantic-pragmatic perspective. We start by briefly discussing the concept of gesture, gesture types, and the controversies surrounding its definition, to then deal with gesture-word combinations, which play a decisive role in the development of semiotic ability, particularly in the process of language acquisition. We then discuss the types of gesture-word combinations described in the literature, followed by different problematic cases overlooked by standard classifications. Finally, we propose a theoretical-methodological strategy that recovers the categories of an observation code devised for word couplings (semantic role duples), which we consider a powerful tool to further investigate the transition from the initial bimodal gesture-word combinations to the first grammatical structures. In order to incorporate all varieties of gesture-word communication, we outline a number of steps leading towards an ecological code of infant semiosis.

The examples presented originate in a case study carried out with the youngest son of the first author. Some observations have already been published (Rodríguez, 2018a, 2021; Rodríguez & Español, 2019) while others are reported for the first time.

7.2 Gesture: Definitions and Controversies

In the 1970s, debates on language acquisition oscillated between committing to a generative grammar and undertaking *preverbal communication* studies. On one side, language was encapsulated as a modular competence; on the other, it was understood as an ability subject to ontogenetic development and social interaction. Studies in preverbal communication incorporated communicative dimensions neglected by the harder linguistic option: suprasegmental phonemes, facial and postural expression, idiosyncratic styles of speech and movement, gestuality, pragmatics in general and the interactions that particularly take place in the child's preverbal stage, etc. (Bullowa, 1979). They also suggested, for the first time, that perlocutionary, illocutionary, and locutionary acts, synchronous in adult communication, could manifest sequentially in the course of development. Thinking that perlocutionary and illocutionary acts could possibly precede the emergence of locution led to further research. It then became evident that the infant's behavior has from the beginning a perlocutionary effect on the adult who interacts with him, and that the intention to communicate with others is present in the gestures and vocalizations

that precede speech (Bates, 1976). In addition, it was observed that there is a gesture constitution process, particularly of deictic gestures, and multiple attempts to explain it. Among the different proposals, it was conceived that gestures result from changes in the action of the infant in an interactive context where the adult interprets, anticipates, and completes this action. It was suggested that changes in infant action are produced by an ontogenetic ritualization process in which action is transform action (abbreviated and/or exaggerated) into behavioral patterns meant to target only a communicative goal (Bates, 1979; Clark, 1978; Español, 2004; Lock, 1978). Thus, infants would seamlessly go from a semiosis of action (a pragmatics responsible for grounding, on the platform of preverbal conventions, the most primitive semantics) to a semiosis involving exercises carried out *ex profeso* for communication purposes. That is to say, babies would pass from actions to gestures (Español, 2004) and then to speech (Capirci et al., 2005; Capirci & Volterra, 2008; Iverson et al., 1994; Volterra et al., 2017).

Before preverbal communication turned gesture into a subject of interest, other lines of research highlighted the importance of communication beyond language, allowing a different semiotic interaction relationship through the body and its diverse expressive possibilities. In the tradition of American cultural anthropology (Efron, 1941/1972; Sapir, 1927/1951), Edward Trager, Edward Hall, and Ray Birdwhistell, among others, laid the foundations for the study of non-verbal communication. They worked on a research program seeking to embrace body movement, voice and utterance inflections, and the use of physical space and objects as components of a broader conception of communicative events (Birdwhistell, 1970; Hall & Trager, 1953). The recognition of the central place of other expressive modalities highlighted a specific aspect that Chomskyan linguistics tended to overlook, namely, social interaction. Indeed, the Chomskyan theory argued that the structure of language was housed in the mind-brain interface and that, in return, external stimuli, reduced to pure verbal stimuli (neglecting socio-emotional dimensions), served only to awaken it. Non-verbal communication, on the other hand, brought face-to-face interactions and the semiotic use of the body to the fore.

While non-verbal communication researchers were keen to point out dissimilarities with verbal expression and to integrate all other communicative forms, they never really delved into the features of gesture. Too unsystematic when compared to linguistic signs, and too structured to be on a par with other communicative-behavioral resources, the gesture did not find its place until later (Kendon, 2004). Non-human primate enculturation experiments (starting with the classic work of the Gardner couple−Gardner & Gardner, 1969), preverbal communication discoveries and the topic of the origin of language (Cartmill et al., 2012; Corballis, 2002; Hewes, 1973; Kelly et al., 2002; Kersken et al., 2019; Liebal & Call, 2012; McNeill, 2012; Pollick & De Waal, 2007) were three main factors in the new attraction generated around gestuality. There was speculation that the early stages of infant communication and the dissimilarities with evolutionarily-close species could ultimately illuminate the phylogenetic debate on language.

The relationship of gesture with action and intentionality becomes evident when we analyze the first communicative infant productions. The same gestural

morphology fulfills different practical functions, subject to the needs and the will to interact expressed by the sender-baby: the pointing gesture is made both to induce someone else to bring the referenced object (proto-imperative gesture) and to share an experience (protodeclarative gesture) (Bates, 1979). These pointing deictic functions were later joined by many initially overlooked others (Kita, 2003; Kovács et al., 2014; Liszkowski et al., 2004) that repeatedly confirmed that the infant points for different cognitive and communicative purposes. Deictic gestures branched out into subtypes: reaching, grasping, showing, giving, hold-up gesture, etc. (constantly raising discussions about the ones genuinely deserving to be considered as such[1]). Along with them, the entity of iconic or representational gestures (McNeill, 1992), arbitrary gestures (called emblems by Ekman & Friesen, 1969), functional or connective gestures (Rodríguez & Español, 2019), palm-up gesture (Cooperrider et al., 2018), *beats* (Ekman & Friesen, 1969; McNeill, 1992), etc., were recognized.[2]

Controversies over gesture categorization stem from the fact that the limit of this concept lacks consensus. Kendon, an undisputed authority in the field, claims that

> [g]iven the nature of gesture as a form of human expression, we cannot establish permanent categories to represent essentially different forms of expressive behaviour. That is, we have to think of the different gesture typologies that have been proposed as provisional working instruments which may be useful within a certain research perspective or interest (Kendon, 2004, p. 107).

The citation grants each single study a prerogative to establish the necessary distinctions according to its objectives, but at the same time the shifts and overlaps between categories within each classification must inevitably resonate on the overall concept of gesture and on, for example, the question of continuity between gestures and sign language (the question of continuity between gestures and sign languages (if there is any binding thread−for a discussion, Goldin-Meadow & Brentari, 2017; Müller, 2018). On the other hand, what about phenomena such as instrumental action or body expression forms that may eventually overlap with gesture? Even accepting a far-reaching definition (among others, Kendon's "visible action as utterance") and taking adult gestures as paradigms, developmental psychology, interested in describing, first, and then in explaining how action becomes more than action, i.e., how meaning arises in behavior, demands a precise boundary delimitation.

According to Kendon and McNeill, gesture is essentially a non-verbal expressive action performed by the body, deliberately carrying some meaning. Following Kendon, gesture is:

> a label for actions that have the features of manifest deliberate expressiveness. They are those actions or those aspects of another's actions that, having these features, tend to be directly perceived as being under the guidance of the observed person's voluntary control

[1] For example, the authors of this chapter object that giving gesture should indeed be taken for a gesture and argue for it to be understood as action (Español, 2004, Rodríguez & Español, 2019).

[2] The indicated bibliography is only for orientation purposes. For a meta-analysis of non-referential gestures in general and their influence on the child's cognitive and narrative abilities, see the recent work of Vilà-Giménez and Prieto (2021).

> and being done for the purposes of expression rather than in the service of some practical aim (Kendon, 2004, p. 15).

'Voluntary' should not be understood as conscious, but as a movement intentionally aimed at decoding. "Usually 'gesture' is not used to refer to those visible bodily expressions of thoughts or feelings that are deemed inadvertent or are regarded as something a person cannot 'help'" (Kendon, 2004, p. 8). Thus, laughter and smiles, the nervousness revealed in a subject by a tense muscular tone might be signs (symptoms), but not gestures. Neither are mannerisms or affectation modes that make the behavior of a certain person recognizable. Likewise, gesture is an "excursion, rather than resulting in any sustained change of position" (Kendon, 2004, p. 12). The gestural movement fulfills a cycle between the onset, its consummation point (stroke), and a final retraction (offset) (Kendon, 1980; McNeill, 1992).

Kendon's interest was certainly far from that of developmental psychology. The task of expanding gesture exploration to its gestational stage falls on the latter. But by shifting perspectives, a controversial focus unfolds. Trevarthen et al. (2011) propose five different types of infant hand gestures: gratuitous (non-deliberate, involuntary), self-regulatory, emotional, indicating (deictic), and symbolic.[3] From this series, the first class not only clashes with the pattern of the signifying or expressive will of gestures highlighted by Kendon, but introduces a second problem: when infant motor skills are still to be governed, is the will to interact reason enough to speak of gestural behavior? Can movement be plainly included in the spectrum of gestures? By slipping into gestuality, as gratuitous gestures, movements induced by stimuli from contingent interactions, we risk omitting a valuable semiotic distinction, namely the one separating signification from communication (Español, 2004; Rodríguez, 2021).

According to Trevarthen and collaborators:

> Gestures are embodied intentional actions that signal three kinds of moving experience, in three different domains of consciousness. They can sense, show and regulate the state of the body of the person who makes them; they can manifest and direct interests to the objective world of physical 'things'; and they can convey the purposes of communication with other persons, or any combination (Trevarthen et al., 2011, p. 4).

Concerning these three dimensions, the second seems to correspond to referential use, the third to communicative interactions, and the first to cases which, following Kendon, could not be considered if unwitting. Between "embodied intentional actions" (intentionality that only implies, from the infant's point of view, the excitement of being able to connect emotionally with others and to be able to interact) and goal intention (which implies communicating meanings to a given receiver and

[3] "*Gratuitous gestures* are self-regulatory, emotional, or object-oriented intentional actions made for the subject. They are not intended to communicate and may be done without any awareness of others, but they can provide 'stolen' psychological information for an observer" (Trevarthen et al., 2011, p. 10). Self-regulatory gestures are also unintended. "Unintended 'regulator' movements effecting comfort or protection form an undercurrent to any deliberate referential communication by speech or gesture" (ibid.). The communicative character of hand movements is not decisive for gesture delimitation according to these authors.

represents the core of Kendon's and McNeill's conception), one has to reflect on what kind of behaviors would better serve the concept 'gesture' not to encompass all forms of "moving experience". In the same text we can read: "Where one draws the line between a gestural movement intended to be communicative to an attentive other, and one that is self-serving, and not communicative at all, remains unclear. For this reason we find it simpler to consider infantile or ancestral gesture not across the problematic conscious-nonconscious divide, but to accept gesture and its communication as inherent attributes of a fundamentally social animal endowed with a single will to move complexly articulated head, face, eyes, hands and feet, and with feeling for the process and quality of the moving" (Trevarthen et al., 2011, p. 10).

> However, the assertion claiming that the line between communicative behavior, particularly gesture, and other bodily movements "remains unclear" can be countered. People do seem to perceive the difference. "[M]ovements made under the guidance of an openly acknowledged intention of convey meaning are directly perceived as such as a consequence of characteristics of the movement's dynamic features which 'compel' the observer to see them in this way" (Kendon, 2004, p. 14). In interactions, a semiotic intuition enables the discrimination between 'pure' behavior and communicative behavior. Other specialists agree with this assessment: "Listeners seem to intuitively distinguish gestures from the stream of other motor actions performed in the course of communication" (Abner et al., 2015, p. 2). "Participants in interaction normally do not doubt the intentionality of other's actions unless there are explicit reasons to do so. In everyday life, movements that look like gestures are the ones that people see and treat as gestures, no matter whether they happen to be performed consciously or unconsciously" (Andrén, 2014, p. 161). This implies that the idea of gesture is intuitively a natural category (not established ex cathedra) distinct from concerning the general movement of the body. Technical definitions, therefore, should follow that nuclear notion.

It is therefore not a choice of the researcher to unite or disunite what the natural concept of meaningful movement represents for ordinary subjects. To unite (as in Trevarthen et al., 2011) the generic expressiveness of the body – which transmits or reveals affective states and forms of vitality (Stern, 2010) by a sort of contagion or direct perception – with the specific gestural expression, made of onset-stroke-offset and which can be split into form and content -a fundamental prescription of intentional semiotic phenomena-, seems to go against the grain of natural distinctions. Even when these distinctions have fuzzy boundaries, they possess a *prototypical* form (in the sense of Rosch, 1973) out of which a suitable concept can be modeled.[4]

While gesture must consummate a certain *form* with which a semantic interpretation should concur, in the expressiveness of the body, in its continuity or temporal extension, the different movements are interwoven, drawing kinetic profiles with unclear beginning and end demarcations. Bodily expression is not made of units and is not reducible to a discrete semantic expression. It was precisely dissimilarities of this kind that made it impractical for Birdwhistell to include gesture in kinesics,

[4] An early vindication of the use of natural categories to elaborate an adequate observational code for the study of preverbal communication is found in Sarriá Sánchez: "[I]t is still necessary to make an effort to establish an operational definition of intentional communication: a definition that enables empirical research, while avoiding artificial restrictions of the concept" (Sarriá Sánchez, 1991, p.363). See also Sarriá and Rivière (1991).

where meaning arises from contextual coordinates and from a type of bodily involvement different from that of pre-delimited communicative units. The imprecise, continuous, expressive movements, certainly segmentable although in insignificant parts, seem to belong to another class of inter-subjective connections needless of interpretation (in a decoding sense). As with the direct perception of emotional states, the signifying side is not split from the signified, feelings are captured and sympathized with rather than decoded.

Most adult gesture studies (but also child) focus on manual activity (McNeill, 1992; Kendon, 2004; Kita, 2003, to cite just a few main works), but admit that other parts of the body can also perform meaningful and communicative movements (the head, the torso, etc.). Understandably, research on the ontogenetic development of semiosis involving very early stages of life, when deliberate manual control is still very limited, demanded that the movement spectrum extended to include the entire corporeality. This wider criterion may have led to the aforementioned disagreements into separating gesture from action and body expression in general.

As Kendon argued, gesture categorization must be tailored to the objective pursued by each particular empirical study, but desirably, the discussion should be healthily aimed at eliminating misunderstandings. At a time when no one fails to recognize the relevance of gestures in both cognitive and communicative processes, particularly in face-to-face situations, nor in debates about the origin of language, it is a sensible claim to ask for precision in their nominal and operational definitions. In developmental psychology there are at least two versions of what may be considered as gesture. The definition outlined by Trevarthen et al. (2011) and the one that (since the time of Bates and other pioneers, closer to Kendon) conceives gestures as well-demarcated, significant in themselves and intentional bodily actions (as corporeal *excursions*). This second approach is consistent with gesture intuition and offers a clear framework for the ontogenetic study of semiosis. Differentiating gesture from expressive movements does not mean denying the existing continuity between these two perfoming types, but emphasizing that the particular conformation of gestural expression emerges through action and movement at specific development moments. We are not trying to make an ontological distinction which for some purposes may be irrelevant. Instead, we aim to differentiate gesture from the broad spectrum of children's kinesthetic configurations to follow its developmental history.[5]

[5] While in studies of communicative musicality, the expressions "sound gesture" and "motor gesture" are used in the same sense as in Trevarthen et al. (2011), and while in previous chapters of this book the phenomena of communicative musicality was approached considering those two expressions, in this chapter we understand "gesture" as an intentional sign, therefore not as if it existed from birth and was a starting point for development, but as an emergent of adult-baby social activity (Español, 2004; Rodríguez, 2021). Similarly, "communication" in this chapter refers to intentional (delivered from a concrete ideational or emotional content), purposeful (voluntary) and signic behavior, in accordance with the tradition of gesture studies in early childhood developmental psychology.

7.3 Bimodal Gesture-Word Combinations in Pregrammatical Communicative Interactions: On some Marginal Cases

The study of the first communicative gestural signs (late 1970s) was followed by the study of gesture-word combinations (1980s and 1990s), whose importance for the ontogenetic research of communicative semiosis has proved to be inflectional. These combinations revealed unsuspected cross-modal cooperation types in child expression, supporting the assumption of a shared cognitive structure between the two sign types against modular models of language processing (Kendon, 1972; Lüke et al., 2020; McNeill, 1992, 2005, 2007, 2017). Whereas for the latter, the characteristics of verbal signs are excluding conditioning factors (even in sign language−see Stokoe, 1960), the non-modular hypothesis advocates for a type of processing that tolerates, on the one hand, arbitrary signs governed by grammatical guidelines (i.e., words), and autonomous, unsystematic, gestaltic signs (gestures) on the other. Such studies have shown that before the child is able to deliver emissions with two signs combined exclusively in the verbal modality, gesture and speech collaborate to produce bimodal couplings with meanings synergistically elaborated from their two component parts (Butcher & Goldin-Meadow, 1993; Capirci et al., 1996). This brought forth the need to review the existing knowledge regarding the holophrastic or first-word-period (according to the standard periodization for native language acquisition studies) and to consider not only verbal signs but also gestural signs, either taken separately or bimodally combined. From there, ontogenetic progress in semiotic ability could be conceived as an effect of the reciprocal underpinning between sign types with different properties that collaborate to gestate complex expressions and lead infant semiosis toward a relative modularization (not modularity – see Karmiloff-Smith, 1992) in face-to-face situations.

Gesture-word combinations began to be studied when they were identified as semiotic compounds of two modalities coordinated in *time* and *meaning*. As for temporal coordination,

> Temporal alignment is also a basic property of co-speech gesture, one that is likely necessary for gesture to be understood by listeners. The stroke of an iconic gesture, for example, is produced in temporal alignment with the linguistic unit whose meaning it iconically represents or supplements, sometimes called the "lexical affiliate" (Abner et al., 2015, p. 4).

In fact, the synchrony between gestural and verbal signs was proposed to be based on a biomechanical arrangement or platform acting at the neurocognitive level prior to any semiotic coupling (Pouw et al., 2020). Rather than thinking of gesture-word synchrony as a process controlled by higher-order psychological faculties, it has been suggested that such timing could derive from the more general coordinated and rhythmic movement of the body. The conjunction between oral expression and movement would thus be a more original behavioral device than infant semantic expression. When infant rhythmic behaviors of infant-adult interactions are unimodal, they decrease between the ages of 9 and 12 months, but if they are bimodal,

they remain stable. When infants synchronously produce rhythmic movements and vocalizations and look toward parents, parents tend to be more responsive and act in a manner more in line with the detected infant behavior (Moreno-Núñez et al., 2021). This shows that bi(multi)modality plays a central role among interaction patterns, generating perlocutionary effects, triggering adult responsiveness (see Chap. 4 in this volume). These unintentional bimodal behaviors could be thought of as a natural system that reorganizes through social interaction, paving the way for communicative development (see model in Chap. 6 in this volume). Between this elementary form of multimodal behavior and communicative multimodal behavior, Murillo et al. (2021) suggested that infant multimodal rhythmic movements involving object shaking can be understood as an intermediate stage between instrumental actions and first communicative gestures. Bimodal rhythmic behavior would be the first interactive device to focus the other's interest on an object. In other words, a rhythmic movement performed while holding an object accompanied by a vocalization would be the first means of getting others to pay attention to the manipulated object and, therefore, a precedent for future joint attention, where objects are referenced through gestures. "The multimodal rhythmic movements can serve as an opportunity for the child to learn how to bring another person's attention toward a reference, which is precisely the goal of deictic gestures" (Murillo et al., 2021, p. 6).

Concerning semantic coordination, the second feature of cross-modal coordination, the differences in entity between gesture and word seem to condition the information that goes into each modality.

> For example, when a speaker utters *the box is near the table*, he or she has encoded in speech two objects (*box* and *table*) and a relation between them (*near*). However, the co-speech gesture produced along with this utterance is likely to encode fine-grained information about the objects (the size of the box and the height of the table) and their relation (how far apart the two are and how they are arranged) that does not appear in speech (Abner et al., 2015, p. 6).

During the holophrastic period, three types of gesture-word combinations can be observed, successively more complex in the amount of information and in the type of connection between the intervening signic units. Butcher and Goldin-Meadow (1993) distinguished redundant (matching) from supplementary (mismatching) combinations. The former simply couple signs carrying the same meaning, the latter compound two signs carrying different information. Redundancy, on the other hand, was subdivided between a full matching combination (saying 'this' and pointing in the direction of the object) and relative matching combination (naming the object along with the pointing gesture). The three types of combination (full matching, relative matching, and mismatching) were labeled as *equivalent, complementary, and supplementary* (e.g., Capobianco et al., 2017; Volterra et al., 2005) or *reinforcing, disambiguating, and supplementary* (Özçalışkan & Goldin-Meadow, 2005).

It has also been observed some sort of communicative progress going from the most elementary combinations, where one sign and the other carry co-referential content and a perfect overlapping of meaning, continuing with co-reference cases but with gesture and word providing different meanings, to finally reach bimodal

structures where signs allude to different contents without overlapping (the supplementary combination embodies some bi-articulated subject-predicate idea: e.g., pointing + 'broken').

At about 20 months, children begin to have a larger lexicon, in types and tokens, making unimodal combinations possible (Iverson et al., 1994), but the detection of earlier bimodal combinations showed that prior to any coupling of words, children have the ability to link two signs using orality and gesturing, thereby linking different contents. Advancing from equivalent to supplementary combinations promised a novel field of study to those who took a continuity approach between semiotics and language, since gesture seemed to occupy in the bimodal combinations the place of precedent with respect to the subsequent second oral modality sign (which would emerge later in the two-word period). In this way, the study of gesture in early childhood had found support for more general hypotheses about gesture and the cognitive processing of speech and gesture taken as a whole together (Volterra et al., 2017).

7.3.1 Some Challenges to the Proto-Propositional Model

The undeniable merits of the sequence of equivalent, complementary and supplementary combinations entail, however, as a counterpart, some disadvantages. On the one hand, the sequence is outlined to the subsequent propositional statements of the child, excluding bimodal expressions unaligned with this horizon; on the other hand, within these combinations, differences in criterion mean that different studies do not account for the same expression in the same way. Not least, non-lexical vocalizations were excluded from the world of combinations, which, if occurring within the interaction, should be conceptualized as communicative. The importance of preverbal vocalizations as predictors of future vocabulary development and language ability in children is well supported by empirical evidence (Karousou, 2003; Karousou & López-Ornat, 2013; Murillo & Belinchón, 2013; Vihman et al., 2009; Werwach et al., 2021). As an important note, and in order to clarify why we support the inclusion of vocalizations and not any bodily behavior (even when it might seem in principle that gestures and bodily expression are related in the same way as words and vocalizations), we will say that (i) like words and gestures, vocalizations have a delimited form, enabling their quantification, (ii) they coincide with gestures and form a combination, and (iii) only those intentionally uttered should be counted. In contrast, as we saw in Sect. 7.2, infant expressive movement is a continuum that does not require muscular control or, much less, any kind of communicative semiosis.

Some Marginal Bimodal Expressions

Certain bimodal gesture-word (or vocalization, or interjection) combinations are not intended to convey a *what*, but rather a *how*. They do not refer to objects or events in the world but are produced as an externalization of emotional states or to dynamize the interaction with others. The youngest son of one of the authors of this text, Bruno, observational subject in a case study (Rodríguez, 2021; Rodríguez & Español, 2016, 2019), provided a paradigm of this kind of combination. On repeated occasions, between the 19th and 29th month, sitting cross-legged on the floor, he expressed exasperation by slapping both hands on the knees and uttering a vehement complaint exclamation ("Hey!") at whatever the partner was doing or proposing. These uses of orality (represented here by the interjection) and gestural modality (represented by the slapping hands) have been unwisely put off, perhaps because they move away from those combinations that follow the subject-predicate format (or make verbal or nominal syntagms) and prelude propositional sentences. This type of combination reveals a sender predisposition, the way in which he/she experiences things subjectively, by joining the noun or designative gesture to an emotional complement (interjection, emphatic gestures, modulated prosody to convey affectation or feeling), or linking a pair of expressive signs (as in the example). Rodríguez and Español (2016) wanted to compensate for what could be considered a category deficit and also counted instances of *expressive* combinations. At certain times, Bruno would smash both fists or both palms on a chair or other surfaces to attract attention and utter interjections of obvious anger; at some other times, he would extend his arms out to the sides in a long, astonished "Ohhhh!"; or he would cry out in protest "No!", eloquently showing both palms as if to display what he was rejecting. The identification of expressive combinations as an independent variety of bimodal combinations resulted in a frequency of 12.5% of the total combinations (Rodríguez & Español, 2016), higher than those of supplementary combinations (9.56%) and slightly below those of complementary combinations (14.6%), indicating their intrinsic importance.[6] Recurrent omissions concerning combinations of this type in children's gesture-speech studies, thinking of combinations as a transition to propositional utterances (from a third-person perspective), neglect the communicative axis of the exchanges (second-person perspective): the constant traffic of non-referential expressions, which is precisely what the chapters of this general work have highlighted.

If the horizon is the proposition (subject-predicate) or the structuring of syntagms, then the expressive combinations, anticipated by dyadic play in primary intersubjectivity, are reduced to an irrelevant form of communication. But strictly speaking, they cannot be considered a random variant, even if we think of the bimodal combinations from the viewpoint of future linguistic communication,

[6] In this study we found no progression in the frequencies of equivalent, complementary and supplementary combinations, but this could be explained by the use of operational definitions partially different from those of Volterra et al., 2005, detailed in Rodríguez (2021).

because in the emphatic gesticulation or in the word or the interjection modulation, we find the precedents for the future use of the modalizers in adult speech.

Some Classificatory Uncertainties

The trichotomy of equivalent, complementary and supplementary combinations is not only affected by what they have dismissed (external issues) but also by a lack of consensus (internal issues). In Volterra et al. (2005) the combinations of pointing + 'it' are not considered as equivalent but as complementary, even when both signs seem to be perfectly co-extensive. According to these authors, the pointing gesture would be accompanied by a second referential sign that does not imply reinforcement or redundancy, but specification: what is the 'this' that has been alluded to by means of the word. But no matter how much the deictic gesture serves to punctuate *which* of the 'these', the one here or the one there, is ultimately the one referred to, the two signs are semantically homologous, both demonstrative, and therefore this value of determinative should not be pre-assigned to the gesture in a universal way and without taking into account the various contextual constraints that could potentially give that place to it. On this point, the most important thing to note is that there are not, in this type of combination, two different meanings relative to a single object, as is the requirement of any complementary combination, but that both signs indicate the same thing in the same way: the near point object ('this') without any mention of an attribute. Another is the case of noun-pointing coupling (distinguished in the study mentioned above), where the word endows the referent with categorical precision ('tree' + deictic gesture) and provides the object with specific features (by placing it, through the noun, in a certain class). Whereas the combination of a pair of deictic signs (gesture and word) would provide only redundant information (unless the child had two 'these' objects in front of him with equal rights and a disambiguating pointing was needed), the noun-deictic combination reveals not only *which* object, but also, through the name, some understanding of its nature or quality.

By counting pointing + 'this' combinations as complementary combinations, together with pointing-'noun' combinations, important differences are lost sight of. These had already been established with the discrimination between matching and mismatching combinations, respectively pointing + demonstrative and pointing + noun (Butcher & Goldin-Meadow, 1993). Cartmill et al. (2012) collected data from 18 children aged 14 to 30 months and concluded that the use of pointing together with the object noun, but not with the demonstrative, predicts the onset of determinative + noun verbalizations (indicating that the distinction between the two variant combinations is relevant). It follows that as long as we do not have a unified classificatory criterion, our predictions will be hampered by limitations for the comparison of results.

7.3.2 *Other Difficulties Concerning Conventional Categories of Bimodal Combinations*

There are several other problems, not unique to complementary communications, related to the resolution given in each case, to what the infant really puts at stake with his or her communication.[7]

> Observation 1: 1; 8 (2). Bruno is sitting in his feeding chair (where he can hardly fit anymore), he finishes eating his meal, looks at the adult and, moving his arm up and down, with his open palm facing the floor, asks: "/Down? /". Used to his routine, he knows that at the end of the meal he is always 'set free' to play.

Since Bruno knows to say 'floor', it follows that the verb he pronounced is indeed a verb; however, some reasonable doubts arise because by assuming that this unconjugated word is a verb, we forget that at this age nouns take precedence and that Bruno could simply be anticipating with his verbalization the daily course of his routine. 'Down' might not be a verb, but a semantically redundant companion to the agentive gesture, an alternative (nominal) designation of 'floor'. If so, we would not be in front of an equivalent but a supplementary combination that uses a gesture to refer to the action of descending and a word to refer to the place of destination (two different meanings distributed between two modalities). One must know the child's vocabulary and word usages, the way he/she calls each thing, and pay close attention to the contextual marks before deciding how a combination is to be considered. The semanticity of the combinations cannot be fixed in advance but from the pragmatics of each concrete situation. Take another case.

> Observation 2: 1; 11 (16). Bruno sees his father fixing his bottle while looking for the lid. The father rhetorically asks aloud, as he is not sure Bruno can understand, "Where is the lid?" Bruno spots it on the corner of the counter, points to it and says "Lid".

Clearly there are two pieces of information, location and object (here named), distributed between two expressive modalities. Initially, everything leads us to assume this combination to be complementary, given the use of co-referential signs. But if we conceive that in this interrogation-response context, pointing may not be a determinative but a place adverb (Bruno would not be indicating "*this* lid'', but "lid *there*''), then the intervention would have to be computed as a supplementary combination (proto-propositional content of the type /the lid (is) there/). The subtlety between the two possible reconstructions depends solely on the scene, and the fact

[7] The observations indicate the age according to the Piagetian nomenclature: 1; 2 (3), meaning one year, two months, three days. Double quotation marks are used for spoken utterances, while single quotation marks are used to indicate the mention of a term or phrase. The meaning of gestural, oral or bimodal utterances is written between pairs of oblique lines /xxx/, with occasional parenthetical /xxx (yyyy) xxx/ which replaces elements missing from the child's utterance and/or necessary for interpretation. The absence of one or more phonemes is indicated by an underscore: x_ xx (e.g.: _anks /thanks/). The absence of a complete syllable is indicated by a hyphen: x-xx, -xxx, xxx- (e.g.: -nana /banana/). Where observation elements from studies by other authors are cited, the writing of the original text is maintained (e.g., the use of capital letters).

that the verbalization follows an interrogation tilts the balance toward a supplementary combination.

This leads us to address certain problems regarding supplementary combinations, the third type of combination in the already mentioned sequence and, from here on, the focus of our journey in this chapter. We will see a case where both signs do not form an equivalent or complementary combination, much less an expressive one, and neither can be accommodated in a supplementary proto-propositional structure.

> Observation 3: 2; 2 (10). Bruno and his father are lying belly-up on the living room floor chatting about anything and everything (in consonance with the speech limitations of B's age). The father asks about an object they are both facing: the light fixture in the ceiling. "Is that light?" Bruno extends his arm pointing and replies "Yes". The father continues, "How do you turn it off?" assuming he would either make a finger gesture to press on the switch or go directly to flick it. Instead, Bruno replies "Like this"/ and closes his eyelids.

Bruno decided to show what happens when the light goes out: it becomes dark, and he externalized it with an iconic (representational) gesture copying the effects of turning off the light. The gesture was not materialized using his hands, albeit both were available, but by the body part Bruno considers most suited from his lying position to express what happens. He uses his eyes, which are usually thought to have all their significant power in the gaze, and uses them to release a gesture that is not a facial expression but, in an unprecedented way, a creative gesture that lets us glimpse her thoughts about turning off the light. He does not answer the question of how to turn off the light, but rather indicates what happens *when* the light goes out.

The response embodies a type of combination disengaged from the two-content scheme, grammatical subject + predicate, or object + attribute. The combination 'like this' + the eyelid gesture is, as such, a challenge to standard classification. The gesture, in perfect concert with the verbalization, in a way continues the verbalization, it is placed after it as its completion. Without this gesture, the verbal expression would remain unfinished, because "like this" demands some specific ending. Although there is no redundancy between gesture and speech, neither is there a proto-propositional format characteristic of the supplementary combination, but rather, *bimodalized*, an adverbial syntagm.

In cases where an element plays a vocative role, a different type of link between signs of bimodal combinations arises. Whether it is a contact movement to draw attention, a noun ("Mom!", "Marco!" -Bruno's big brother), or an interjection ("Hey!"), the vocative also lends itself to the type of synchronic combination that does not lead to the verbal proposition to come.

> Observation 4: 1; 8 (28). The mother asks Bruno to call his brother because dinner is ready. Bruno goes into the room and calls: "_a_co /Marco/, am! /food/" pointing in the kitchen direction. In this multi-gesture-word combination of three elements, the pointing gesture coincides with the first word extending to the enunciation of the second.

It is not necessary to specify whether the "Am!" refers to the food or to the physical space of the kitchen; analytically, there would be either three concerted units sharing two roles, vocative-object (the food-kitchen, indicated and at the same time

verbalized), or playing three distinct roles, one for each sign: vocative-object-locative. In one or the other alternative, we have a combination of gesture and vocative very different from what is prescribed with respect to supplementarity.

We must distinguish the previous vocative sign example from other cases where the child points to the adult, assigning him an agent role, to help him achieve what he cannot do by his own means.

> Observation 5: 1; 8 (30). Bruno wants a toy placed on a shelf far from reach. Upon seeing it, he exclaims "Ah!" and makes a reaching gesture (it is a gesture, not an action, as he knows that extending his arm will be useless). He then says "Car" and points to his father.

There are two possible interpretations. Either it should be assumed that Bruno has absurdly called the father by the name of the object, or this gesture fulfills a different function here: it requires the adult to perform the necessary action in relation to the object he has named simultaneously, suggesting /you (are the one in charge of providing me with the car)/. In this case, the gesture-word matrix does not operate descriptively but under the conative function of language. The gesture is not a vocative, because Bruno knew that the interlocutor was attentive to his behavior and therefore did not need to use a sign. Under this assumption, Bruno used the minimal amount of signs needed: one for the agent of the action, another for the object or end. Unlike what is seen in observation 4, there is no vocative here; the combination is a standard example of a supplementary combination. Of this one, however, all its limitations have not yet been revealed.

> Observation 6: 1; 6 (7). Bruno wakes up very early and goes to the bedroom where his parents are sleeping. He climbs into their bed and receives this admonition: "Marco is sleeping in his bed. You have to sleep in your bed". To which he responds by pointing to the parents' bed on which he is already settled, and extending the pointing gesture along two words: "No, Mom".

It is not easy to decide whether Bruno is indicating that he wants to sleep next to his mother or in his mother's bed, but it is absurd to assume that by pronouncing "Mom" and pointing to the bed Bruno is mixing the concepts /mother/ and /bed/. "Mom" fulfills one of these functions: it is genitive, with respect to bed (/mom's bed/), or it is a circumstantial adverb for companionship (/sleeping in bed with mom/). 'Bed' has been tacit in either alternative. In terms of the combination type: how should we compute examples where one of the signs lacks a reference, is not in a designative role, and is instead *negatively* linked to something previously said by the interlocutor? Bruno refuses what has been prescribed to him and then retorts by a counter-proposal: to stay in the bed to which he has moved.

Indeed, the words revolve around a single factor (whether or not to remain in the bed of the parents): the first seems to challenge the command to return to his bed, the second, on the other hand, reveals his intentions. If supplementary combinations are understood as a chrono-semantic cooperation of signs differing in content, then the signs in observation 6, albeit without subject-predicate or object-attribute structuring, should be considered as such. However, the presence of signs with different significance is not reason enough to claim the existence of supplementarity. In this observation there is a chrono-semantic combination of two distinct signs: 'no' is

linked, by sense, to the second affirmative part of the enunciation, and the two parts are colligated by the gesture. The non-relevance of the example to the category of supplementary combination should not prevent it from counting as a combination, because the items of temporal overlapping and meaning are satisfied. The pointing gesture begins at the same time as the verbal expression of negation, although what is pointed out is not the same as what is denied. If we consider that later in adult speech we find juxtapositions of this kind /no X; yes Y/, there is no reason to dismiss this type of intervention.

Consequently, the challenge is posed to consider in an exclusive way that combinations can only exist if the signs are ordered according to the propositional model.

7.3.3 Negation-Inclusive Combinations

The following examples (7–12) allow us to focus on combinations that include negation:

> Observation 7: 1; 9 (28). The father asks Bruno where he left the cookie he was eating. Bruno answers "No" and points to his mouth, indicating that he ate it.

The locative pointing means/it went this way/, while "No" means the object no longer exists or is no longer available (/I ate it/). The combination indicates that there is nothing left while it marks the final destination of the cookie. The meanings of gesture and word are neither dissociated nor fully redundant, so it cannot be a case of equivalent combination. It is not a complementary combination either, because negation and pointing cannot be interpreted as divorced (like the cookie did not disappear from *there*), but quite the contrary. Finally, as we have already seen, the supplementary combination definition requires the expression to be of the subject-predicate or object-attribute type, which is not the case.

> Observation 8: 1; 7 (13). Bruno and his mother interact while sitting on the floor. Bruno has a toy dinosaur and places it in front of her in a threatening attitude, saying "Am, am" (/I eat you, I eat you/). She asks for the dinosaur and extends her hand for it. He agrees and hands it over. She puts the dinosaur on her chest and asks "Is this one mine?". Bruno answers no. She replies, "Whose is it?" Bruno is puzzled. The mother asks again. As the child does not respond, she changes the strategy. She presses the doll against her chest, to make clear that it is hers, and changes the question: "Do you have one for you?" Bruno understands that she has taken away his dinosaur and reacts by repeating "No", bringing his arm towards the toy and grasping it firmly with his hand. He does not want to snatch the dinosaur but to show his displeasure by repeating "No" along with the action of opening and closing his hand. Bruno continues to move his lips as if to say 'no' but now inhibits any kind of verbalization: the babbling continues so that the idea of denial accompanies the hand gesture till the end. If we follow orality, it states that Bruno does not want the mother to keep his dinosaur, while gestuality, on the contrary, insistently demands it. In a linguistic translation, Bruno would be proposing /don't have it you, have it me/. The 'no' of the oral modality, outside this situation, would be antagonistic to the request in charge of the hand, but the contextual coordinates offer the necessary key to correctly interpret the expression.
>
> Observation 9: 1; 8 (15). The family is vacationing at the beach. All their members, parents and older brother, are either lying down or sitting under the umbrella. Bruno

discovers a bag with a package of cookies hanging from the rods of the umbrella. He raises his arm in a reaching gesture and says "No". Prima facie, the marked gesture-word synchrony raises a contradictory sensation, like wanting with the hand and not wanting with the verbal expression. The right sense of this communication would be: /I cannot reach the cookies, I want the cookies/ /do not reach, do want/.

Observation 10: 1; 9 (13). Adults are chatting at the house of his grandparents, while Bruno picks up a pack of cookies from the counter and attempts to tear out the wrap. Vainly, he scratches with his finger for a few minutes until he realizes it is useless and turns to his mother. He stands in front of her for attention and rhythmically repeats the movement he made with his finger (now just insinuated and lacking the force he put in before, since he only wants to show his mom). He insistently says "No", one 'no' for each gesture (representational type) showing what he was unable to do.

In the first observation of the latter two (8), 'No' concerns a question, and the gesture extends like a claim, an ownership assertion /*not* you, *yes* me/. In the second observation (9), the expression follows the same structure: 'no' goes against the intentions reflected in the gesture, although the opposition /no – yes/ is also informative, like in observation 7: /*I can't* (reach) the cookies/. It would be wrong to think that "no' 'and the reaching gesture contradict or oppose each other. Underlying this opposition lies the same idea manifested in conflicting signs. For observations 8 and 9, what at first seems incoherent (yes and no by means of two simultaneous signs), bimodally reveals the deep semantic connection between both components. In the third observation (10), two signs are used, negation + X, which follow the normal pattern of a negative statement. Both signs are semantically concerted within the same bimodal sentence: a functional element, negation itself, affects a content (/I *can not* (open) the package/). Thus, we can see how the pre-grammatical cognitive-communicative structure captures, through bimodal expressions and using signs executed in perfect simultaneity, the same type of message that language, necessarily successive, must unroll and arrange on a linear-unimodal axis.

Observation 11: 1; 10 (28). His mother insists Bruno to eat a cookie he is carrying around with him. He refuses every time and decides to bring it close to her mouth so she eats it instead. He says: "No, _ummy" /No, yummy/.

Here, 'yummy' refers to the cookie. It is not that Bruno is not used to say "_ookie" /cookie/; it is just that in this case, the adjective 'yummy' seems to take the place of the usual noun (or else, it would be the reason why she should eat the cookie: because it is yummy). Both verbal sign interpretations represent a new case of /not me; yes you (eat this yummy)/.

Observations 6–11 exhibit different negation forms in the bimodal combination. Observation 6 is about two signs, negation and noun, each behind different targets in perfect synchronicity. In observation 7 the word "no" is allocated according to what is gestured: /there is nothing left (because) the cookie went this way/): a combination where the second component specifies or justifies the content of that negation. Observations 8 and 9 offer a juxtaposition which, by the orientation of each sign towards different fronts, reveals a different type of combination: no X / yes Y, delivered bimodally in parallel. Finally, observation 10 provides a typical case very necessary for the grammatical organization of 'no': 'no' + a complement ('no' *with*

respect to X, not ➔ X), which here corresponds to /I cannot perform what I am gesturing (the opening of the package)/.

In cases 8, 9, and 11 it is evident that bimodality, in addition to admitting the coupling of two convergent signs on the same content, or divergent signs on two aspects of an object or event, manages to link, by overlapping them in synchrony, pairs of elements that eventually go in opposite directions (no/yes at the same time). This kind of opposition foreshadows the subsequent assembly of antagonistic juxtaposed propositions. Unlike the equivalent, complementary and supplementary combinations, these combinations do not only materialize a sign compound, but what in linguistic terms would be even a broader statement, an expression with partial antithetical elements, the germinal structure of the future molecular proposition. Indeed, while the supplementary combination is centered on the subject-predicate or object-attribute formats, precursors of any standard major sentence, and therefore constitutes the bimodal structure that will later give rise to the *atomic* proposition, observations 8, 9, and 11 allow us to think that there are cases of combinations of a different nature, which convey two independent ideas and thus anticipate the *molecular* proposition. It seems then legitimate to assume the existence of molecularity in bimodal combinations. These are cases of communicative expressions that integrate contrary ideas by connecting signs of different extraction. If we understand that combinations of the pointing + X type correspond to /X mine/ or /X mom's/, and that later will come to be verbalized as an atomic proposition: 'X (is) mine' or 'X (is) mom's', in the same way two signs with opposite senses, each one in itself a holophrase and both emitted in a synchronized way, can contain *in nuce* two propositions juxtaposed in a unit and, as a consequence, bimodally preannounce that which will be the verbalizations of molecular class.

It is not surprising that two successive signs show in a holophrastic way, each one by itself, two proto-propositions of opposite tenor, one negated and the other affirmed. What is striking is that the option of denying and affirming different things simultaneously can already be found in bimodal compositions. If we were not dealing with pregrammatical children but with competent adult speakers, in order to convey a similar content with language they would have to use two sentences or words of opposite sense and pronounced successively.

Observation 12: 2; 2 (29). Bruno brought back from kindergarten a little tray containing his first craft, some seed and noodle-encrusted dough. With his mother, they are now in the kitchen commenting on this initiatory sculptural work. Pointing towards an element, he says: "This _ookie" /this (is a) cookie/. His mother says: "It's not a cookie, it's a noodle". B: "¿_oodle?" Mother: "Yes, and you don't eat it". B: "Eat it?" (replacing the negative component in the mother's statement with a nod). M: "No, you can't eat it hard like that". B: "That hard?" (again nodding his head in denial).

Observation 13: 2; 4 (10). Bruno is playing with her babysitter R. They are sitting on the floor with some toys scattered between them. Bruno wants to put a battery in the Schrek doll, as he often does with his other dolls. R tells him it is not possible: "It's not for Schrek, this". And Bruno docilely replies: "_uek this and repeats the negation in the gestural modality, closing his communication. He extends the index finger and moves it, on the same plane, from one side to the other. As in the previous observation, a segment of his adult

partner's utterance has been converted into a sign of another type and moved from the beginning to the end of the message.

In observations 12 and 13, Bruno transforms the negation expressed verbally by the adult into the gestural modality. For convenience, he would have been expected to reproduce the input as an output on the same channel. But for some reason the whole linguistic message was disassembled and split into two channels. Only a part of the original remains in the oral form. The basic structural bimodality is recognizable when we notice that a register can relieve the other or delegate a certain content to it. Bimodality is not only in the combinations produced, it also reworks perceived communicative materials and manipulates them at convenience to formulate the message in an altered way. If it would seem easier to repeat, in a normal situation, what has just been heard, then it is worth questioning why Bruno would have opted for this conversion of the verbal 'no' to the equivalent gesture. Perhaps a limit in the possibility of repeating a sequence of six word units diverted a sign to another channel, with the advantage of alleviating the processing of a long sentence. For whatever reasons, it is clear that there is an open flow between the two channels of expression and reception.

7.3.4 Double Gestures and Compound Gestures

Generally speaking, considering all the preceding observations, we note that some bimodal expressions do not fit the parameters of normally studied combinations. The collection of borderline examples or examples poorly adjusted to existing patterns demands a categorization that enables following, at a shorter distance, the transformation of gesture-word (or vocalization) combinations into word-word couplings, first by juxtaposition, then according to morpho-syntax rules. In other words, the three pioneer categories addressing bimodality in the pregrammatical child (equivalent, complementary and supplementary combinations) have overlooked a wide range of other communications, ignoring the fact that these other combinations persist later in the competent adult speaker, either unimodally or bimodally. They should be considered and accounted for in the interest of investigating the passage from pregrammatical to grammatical communicative forms.

We cannot imagine that examples such as the observations reported here have not crossed the path of other researchers, so we must assume they were relegated to the 'other cases' bag up to the purposes of each specific study. It is impossible to imagine that examples such as the observations reported here have not crossed the path of other researchers, so it must be understood that they have been relegated to the bag of case-other according to the objectives of each specific study. If we add to this the fact that some experts have deliberately cut out not only vocalizations but also non-manual gestures or gestures involving manipulation of objects (Butcher & Goldin-Meadow, 1993), or have systematically discarded emblem gestures

(McNeill, 1992), the capital of expressive resources is reduced to an important degree.

On the other hand, in addition to the difficulties posed in gathering the full spectrum of infant bimodal expressions within the gesture-word combination categories, coupled with the classificatory limitations resulting from the neglect or proscription of the group of vocalizations, there are still certain varieties of gestures lacking proper identification and cataloging. When Bruno says (undated observation, around 2 years of age) he has finished his bite ("No mo_"? /There is no more/) and in one sole action opens his mouth and points inside, two crystal clear gestures concur to help better visualize the idea in question. The open mouth is a gesture (of the non-manual subtype) requiring no reinforcement, since it is already verbalized ("No mo_"); however, the pointing gesture eliminates any possible doubt: it is in this cavity, where food goes, that there is nothing left. The open mouth summons the hand to round off a more finished meaning between them. The gesture is *double,* comprising two movements, each in itself a gesture in its own right, producing together one single meaning. If we wanted to retranslate this reinforcement between two gestures to the oral modality, we would be forced to think of expressions that would emphasize some content in order to make it doubly visible, expressions that would necessarily arrange the signs involved serially. In the gestural modality, which unlike speech chain is non-sequential, the finger points toward the oral void, specially deployed so that there is a place to point. The two gestures are arranged around the idea to be conveyed.

But this *double gesture* category thus identified must be distinguished from any unimodal combinatorial expression of gestures. If in the double gesture two gestural signs overlap in one and the same meaning (the food already swallowed), and if therefore it represents an example of complementary combination (or even, perhaps, equivalent combination), carrying out two different gestures with different semantic charges in order to jointly generate another meaning simply implies a unimodal supplementarity. As an example (new undated observation, around the age of 2), Bruno brings an open hand to his chest and wordlessly points with the other hand to the exact place where he asks to be lifted. By coordinating both movements simultaneously, when the first hand rests on his body, the other extends to the maximum, and the meaning /me there/, conveyed in complete silence, is clearly understood. Insofar these two gestures−one referring to his person and the other to a destination−are temporally and semantically linked, we must accept they are it must be accepted that they are conceived to share between them an integral message. This is a second example where Bruno's older brother, Marco, provides a brief anecdote at the beginning of a vacation trip (third undated case and several years earlier, since Marco was approximately two and a half years old at the time). Upon arriving at the rented apartment, while the parents were packing, Marco started running around the rooms; when he had gone through them all, he sits down next to where the parents were unpacking. When he hears the father's question about the number of toilets, Marco hurriedly intervenes and says: "Two", pointing in two different directions with each of his hands. Each deictic gesture points to an individual referent in different locations and independently, providing information absent in

the verbal expression. This is not a double gesture. Each hand made a particular indication. Each hand makes a particular indication. However, an association is noticed between both gestures, inasmuch as both hands perform their individual action in unison and are integrated, together with the verbalization, in a single unified response (not only were two baths reported, but also their respective locations). In contrast with the double gesture of convergent meaning, this type of association could perhaps be called, bearing in mind the genuine autonomy of each gestural unit, but also the synchronous performance, a *compound* gesture.

The purpose of this section was to demonstrate that combinations of equivalent, complementary and supplementary nature, which of course record what they are designed to record (even if there are conflicting cases), do not fully comprehend every bimodal combination variety. But at the same time, if we are to look for connections between gesture-word (or gesture-vocalization) combinations and the forms of the grammatical speech, we must probe the semantics inside all combinations. If this semantics will normally derive in syntax (and if syntax is somehow conditioned by semantic categories -see next section), then investigating *what babies and very young children do* with gestures and first words, what roles are assigned to them in their bimodal communications of two or more signs, becomes imperative.

If deictic gestures can fulfill, much like personal pronouns, the function of the object or the executor (if it can show us the ball or who is kicking it), and if it can be coupled to a word and share or distribute with it the two contents to be expressed, the study of transitional semiotics between the first uni- and bimodal ungrammatical combinations and the seminal expressions with morphosyntax must address this question. Something else is needed besides distinguishing deictic from arbitrary or representational gestures, or combinations between gestures and words, because gesture and word exchange their roles when they participate in different expressions. If one single gesture is used for different purposes, if a verbal or word deictic serves to indicate *who and what*, it is crucial to inquire, together with the support of signification (support, vehicle, representamen, or signifier) the relational semantics of the combinations or the roles underlying the construction of couplings.

7.4 From Bimodal Combinations to Grammatical Structures: Looking for a Missing Link

The typology of bimodal combinations in early childhood has hardly changed since the pioneering texts. The categories of equivalent, complementary and supplementary combination, of undisputed fecundity, were elaborated from and for a transitional stage, leaving to linguistic research the relay of word-word forms in their march to grammar. A different tool for analyzing gesture-word (or vocalization) and word-only combinations is offered precisely from linguistics and its usefulness has not, to our knowledge, been exploited in the field of gesture or the field of children's

bimodal combinations. In the 1970s, Charles Fillmore's case grammar highlighted verb-noun nexuses (agentive, instrumental, dative, factitive, locative, and objective cases – Fillmore, 1968). These cases do not follow the subject-predicate ordering and are autonomous with respect to the syntactic level. The 'objective' case noun can take the place of subject or direct object within the sentence. In examples such as 'The door is open', or 'Peter unlocked the door', 'door' embodies the notion of object in these two different forms. Likewise, 'Peter' exemplifies an agentive case, but similarly it can function as an objective case if we say 'Laura kisses Peter'. Fillmore's work revealed that semantics plays a determining role in grammar, since it presupposes and determines sentence structuring by the number and type of valences implied by a given action. *Semantic roles*, or cases, exposed that the plexus of transformations theorized within generative grammar was structurally conditioned by thematic links. Verbs are normally transitive or not (excluding counterexamples) by themselves, before they receive adjustments in a certain phrase.

7.4.1 Role Semantics or Thematic Pairs

Also at the beginning of that decade, Bloom (1970), Schlesinger (1971), and Bowerman (1973), following closely the first combined child verbalizations, postulated relations between types of words not coming from verb semantics but from the word pairs most regularly found in their observations. If Fillmore's approach was essentially theoretical, these researchers, devoted to the description of the beginnings of language, proceeded by collecting examples. The tandems recorded in effective infant communications reflected the most frequent word combinations and enabled the establishment of combinatorial types. In this context, Roger Brown (1973), bearing in mind those other works but based on his own casuistry, proposed a classification into three referential classes (*nomination, recurrence, and nonexistence*) and eight "*prevalent semantic relations"* (Brown, 1973, pp. 172ss.): agent-action, action-object, agent-object, action-locative, entity-locative, possessor-possession, entity-attributive, demonstrative-entity. As he clarifies, "[d]escription in terms of a set of prevalent semantic relations may be little more than a technique of data reduction, [but] [e]ven as a reduction technique it has certain value. It reveals the fact that, *Stage I* utterances, in all the languages for which studies exist, concentrate on the same set of meanings" (Brown, 1973, p. 173).[8] These categories accounted for 70% of the total expressions of Stage I samples (Brown, 1973, p. 178). Less frequent cases and unclassifiable cases were left out. Cases with salutations or vocatives, which we have vindicated here, were relegated to the large residual category 'others', given their irrelevance to grammar.

[8] Stage I covers *Mean Length of Utterance* ranging from 1.75 to 5 morphemes involving two-word constructions characterized as primitive sentences.

The referential categories mentioned above only contemplate single-word utterances. When the latter was linked to a second element (demonstrative), the utterance was then considered in the demonstrative-entity relational category. Brown warned very clearly that this procedure involved an artificial maneuver, because either way it meant that "names and manifest referents are linked and nothing else is expressed" (1973, p. 189). He pointed out that in those instances the name was usually escorted by a pointing gesture. It is not necessary to stress that gesture did not fall within the interests of Brown, yet it is clear that from a multimodal appreciation, but it is clear that from a multimodal appreciation it is possible to challenge such a taxative assertion as the one that, between a noun and a demonstrative, "nothing else is expressed", since the semantic roles distributed between gesture and word (synergistically combined) perform different functions. As we have already seen, when a pointing is paired with a noun, the latter adds a conceptual charge to the deixis and the former indicates in which concrete entity the universal feature is materialized (notwithstanding the fact that in other circumstances the pointing gesture can indicate an entity, or a place, and the word can name a different entity as a supplement of the mentioned object).

Among the remaining referential categories proposed by Brown, the recurrence category comprises the samples underlying the idea '*more than X'*. Finally, the category of nonexistence concerns the primary mode of negation, prior to the negation form of X ('not X'). Although the referential categories can be expressed with more than one verbal sign, they are conceived as representative elocutions of a fact or object in both cases unitary. Relational categories, on the other hand, correspond to verbal combination structures. Here is the detail of relational categories taken from Brown's casuistry (1973, pp. 193–209).

1. *agent-action*: 'car go', 'mommy fix'.
2. *action-object*: 'hit ball',
3. *agent-object*: 'Mommy pumpkin',
4. *action-locative*: 'put floor',
5. *entity-locative*: 'baby table' (baby sitting at the table),
6. *possessor-possession*: 'Daddy chair",
7. *entity-attributive*: 'dirty my hands",
8. *demonstrative-entity*: 'this ball' (exemplified from considerations on Brown, 1973, p. 189).

As relations are semantically defined disregarding syntax, word duples do not embody ordered pairs and the cases can be displayed with an inverse arrangement in the expression (see duple 7, entity-attribute). They are binomials, which is why seriality, so important in the English language, is not taken into account. We technically distinguish pairs from duples to speak, respectively, about the two-word strings in Brown's classification and the uni- or bimodal sign structures that can be named and serialized according to the number of components: duples, triples, etc.

Infant semiosis research, at the gesture-word combination level, can clearly benefit from a thematic relationship or semantic role categorization once we are aware the child is able to link, as a supplementary combination, two different contents in

different modalities. In other words, what Brown and other experts isolated in the 1970s and applied to the first word combinations can be taken upstream to explore the semantic roles between gesture and word. Since combinations are not fixed tandems, and the different word signs (e.g., a noun) do not invariably have the same function when coupled with the same gesture (combining noun and gestural deictic could involve the agent-object and also possessor-possession duple), addressing these relations in bimodality would be a way to take infant expressive combination studies forward. This list of pairs will allow us to know which types of bimodal structures prevail among the gesture-word combinations, which components of the child's situational picture in each case are signified, which are postponed as part of the backdrop that the interlocutors must keep in mind to understand the expression, which cross-cultural regularities and which differences in individual performance can be traced, as well as whether – for certain uses – the agent-object duple is preferred with a gesture for the first role, while in other circumstances a word is preferred. Research on the ontogenetic development of semiosis owes itself a progression chart showing how duples (triples, tetrads, complexes of sign units) are produced in infant communication according to what the child wants to (and can) transmit.

The use of semantic roles beyond their original application matrix (speech) offers a rich horizon of opportunities for inspecting the passage from bimodality to verbal modularization (with no encapsulation). At the time, the investigation of semantic roles had reached an unsurpassable instance, but today we can move forward, indeed backward, beyond the threshold of word combinations to the dawn of expressive gesture-word (or vocalization) combinations, and describe the cognitive-functional restructurings that are taking place. Under this uniform code of semantic roles, gestuality and verbal expression disappear as separate formats in favor of a fundamental semanticity. The research of bimodal combinations will finally be able to transcend the limit of supplementary combinations, which have operated as a *non plus ultra* frontier for failing to contemplate sign relationships from a thematic angle. For its part, research on language acquisition will have gained access to the step prior to its unimodal couplings, to the origins of *complex* expressions, surpassing a historical cut-off line. If the work of Brown and others brought children's speech to the minimal word-word couplings, the thematic pairs that we have inherited from them offer an extension of their work in the space of early bimodality.

Unlike in Fillmore, semantic pairs do not cognitively respond to the structure of language, but are rather factual-perceptual or *performative* organizers of infant environmental experiences. For the establishment of these pairs (which children intuitively use to make themselves understood), the body contributes to the expression in an unavoidable way with its versatility. For the establishment of these pairs (which children intuitively use to make themselves understood), the body contributes to expression in an unavoidable way with its versatility. Before executing one sign or another, be it a deictic or representational gesture, it creates for itself, through movement, the categories of activity, object, spatiality, agency, etc., the content of which will then need to be communicated. The connectors between the self and others, objects and the immediate environment originate in the body, ground zero of

cognitive possibilities, and fundamental scheme of the word according to its capacities. Signs are built on this inaugural outline of the world, which at the beginning are just detachments of action in space, with or without objects and within the framework of motive-affective exchanges with peers. Therefore, semantic roles derive from the relationship with things and others where infants are occasionally agents, objects, means, or recipients, a game of shifting values where each element – infant and fellows, each surrounding thing – changes along with the rest. A game the child must first of all come to understand in the domain of immediate empirical reality to be able, later on, to trim and collect in the corresponding signic format.

Brownean semantic roles can be effortlessly illustrated with bimodal combination examples of our own:

1. *agent-action*: 1; 9 (5). A bang is heard in the next room. Bruno appears to inform the parents, saying "I" and making a representational gesture with an open palm (a small blow to the head) announces that he hit the part he touches (the blow is understood by the rapid and explosive movement he touches his head with).
2. *action-object*: 2; 0 (10). He says "_reak /break/" and with his open hand shows the tower of cubes he just knocked down.
3. *agent-object*: observation 11, *supra*. Adult-directed pointing and "Car" verbalization.
4. *action-locative*: 1; 9 (0). He is sitting on the floor with his parents, stands up and points toward his room, while verbalizing "Go".
5. *entity-locative*: 1; 9 (8) Bruno finishes playing in the grandparent's garden and is sweaty. His father says, "Shall we go wash your face?" Bruno repeats "_ace?" and points in the bathroom direction.
6. *possessor-possession*: 1; 9 (0). Bruno is in his bedroom with his toys. His father brings some others left in the living room. Bruno recognizes a particular one and before his father puts it away, he points to it and says "Me" /= mine/.
7. *entity-attributive*: (in different occasions) Bruno points to a beloved object and adds an overgeneralized positive qualifier ("_ute" /cute/) or points to an ugly or feared object and adds an overgeneralized negative qualifier, e.g., "Bad" pointing to a dog that barks loudly.
8. *demonstrative-entity*: 1; 10 (6). Playing dolls with the babysitter, she asks him which doll is a new present. Bruno has that doll stuck through the elastic of his pants, where he slipped it in as part of his play. Since both his hands are busy, he responds by saying "This" and sticks out his belly in a gesture. To keep the other two toys in place, he exhibits the doll sticking out of his waistband in a different way and thus complements, combinatorially, the verbal expression.

Tracking what goes on inside the combinations is the key to unlock the division between the initial semantic-pragmatic bimodality and the subsequent unimodal verbality. If semiotic intention is subtended by cognitive links of the type *what-who-where*, or *who-what thing-where-for whom* (among others), which the child must be able to reduce to pairs, understanding how the reduction takes place, which context components the child selects to put into signs (initially limited by a very restricted lexicon but with sufficient perspicacity to cut out certain elements from the

perceptual picture, ponder and *signify them*), how this selection and communication capacity matures, requires that both bimodal and unimodal combinations can be accommodated in an instrument that is capable of both categorizing and quantifying. Semantic pairs allow the integration of bi- and unimodal combinations in a promising way to fill the gap that still separates gesture from language, pregrammatical couplings from correct speech structures. Thus, the categorization by roles and thematic pairs emerges as the missing piece connecting two halves of a continuum capable of showing, with adequate field research, the transition from the first deliberate semiotic interaction to the modularization of language (a point of arrival, it must be said, unfailingly relative, because face-to-face interactions continue to intermingle gestures with words and other forms of expressivity).

7.4.2 Formal Models and Semantic Roles: Linking an Old Idea with New Observations

While coding through semantic roles was administered to the first words in the 1970s and seems highly fertile for bimodal combinations, a different tool has been used more recently to address supplementarity cases in a particular way. In the style of the cognitive modeling profusely proposed in the 1970s and 1980s, multi-word speech and gesture-speech combinations were separated according to their internal structure into predicates and arguments, distinguishing predicate-predicate tandems ('I like it' + eat [iconic]), argument-argument ('Bike' + helmet [point]), and predicate-argument ('Hair' + wash [iconic]) (Özçalışkan & Goldin-Meadow, 2005; see also Özçalışkan & Goldin-Meadow, 2009, 2010). Each predicate has, semantically, essential links according to the language structure and what the utterance is intended to convey. There are unary (being tall ➔ someone), binary (hitting ➔ someone + something), ternary and n-ary predicates that the child understands from the situation he/she wants to describe. This diagramming allows us to follow how these valences are revealed in the child's expression. Each type of gesture had a predetermined role in these structures:

> Supplementary gesture-speech combinations and multi-word combinations were categorized into three types according to the types of semantic elements conveyed: (1) multiple arguments without a predicate, (2) a predicate with at least one argument, and (3) multiple predicates with or without arguments. Each of these three categories was further coded in terms of the number of arguments explicitly produced (…). Deictic gestures were assumed to convey arguments (e.g., point at a ball = ball, hold-up a cup = cup), and iconic gestures were assumed to convey predicates (e.g., moving hand to mouth repeatedly = EAT). Conventional gestures could also convey predicate meanings (e.g., extending an open palm next to a desired object = GIVE), but not arguments (Özçalışkan & Goldin-Meadow, 2009, p. 4).

It is not clearly understood why an iconic gesture, such as sticking out the tongue and panting *like a dog*, should be a predicate instead of a simple argument (a nominative allusion to the dog as such), nor why a deictic *asking* gesture qualifies as

conventional, nonetheless considering supplementary combination in this way makes for an interesting alternative. At the same time, the advantages should be weighed against the decision of starting from the predicate-argument split. This binomial sorts out combinations based on a logical distinction that children will most likely fail to make in their use of signs. Their communicative skills are conditioned by their affective-interactive and sensory-motor experiences. We should not presuppose that children are able to make this separation. Rather, it seems to be a superimposed disciplining of their communications from adult structuring. It might be adequate for the language-thought interface, but not the best policy to follow in the footsteps of semiotic evolution.

Even if this coding contemplates cases of combinations of two arguments without a predicate, the very idea of argument presupposes as its counterpart a predicate without which nothing deserves the name of argument. To identify an expression having only two arguments, such as 'boy' + ball (with this second object equally verbalized or indicated with a gesture), would also compromise the assumption of an absent predicate (/touch/, /move/, /kick/), which, even if *imagined* to exist in the cognitive processing of the sender-child, can be ignored for the benefit of the doubt and without collateral damage. Under the logical perspective adopted, even an unexpressed predicate must be assumed to be latently co-present. But strictly speaking, the argument *may* never be incorporated into the child's mental schema and only satisfy the requirements for applying a tripartite (or bipartite, if a single argument corresponds) structure taken from grammar functionalization. The strategy is admissible as an abstract elaboration, devised a posteriori and projected onto the expression of the child, but may imply an unnecessary modeling. In observance of the simplicity principle, a more basic reading of concrete facts can be chosen, like the semantic role proposal. Figure 7.1 illustrates the conceptual difference intended to be emphasized.

On the other hand, the predicate-argument model entails another limitation. It can be analyzed in a specific example:

> Observation 14: 1; 10 (14). It is lunchtime. Bruno and his father go to the bathroom to wash their hands. The father opens the faucet and Bruno extends his hand to reach the soap, while verbalizing: "Hand". Both signs refer to both objects, object-instrument and object-purpose, respectively the soap and the hand. Both arguments are not arguments sensu stricto lacking a predicate, but referents colligated by an intention and a foreknowledge of routines and instrumental actions.

Certainly, the idea of washing hands is behind the two referents of the example, but if we want to study concrete expressions, we must concentrate on what the child has decided to – and succeeded in – putting into signs. Strict adherence to the actual expression has the benefit of not unintentionally introducing the very thing that is to be sought, or what is expected to appear at some point on the way to the propositional form. The predicate-argument pattern can describe sentence structures and distinguish between logical and grammatical subjects, but infant cognition at this early age, as revealed in the inaugural symbolic combinations, requires no more sophistication than that proposed in the Brown-style bicomponential scheme.

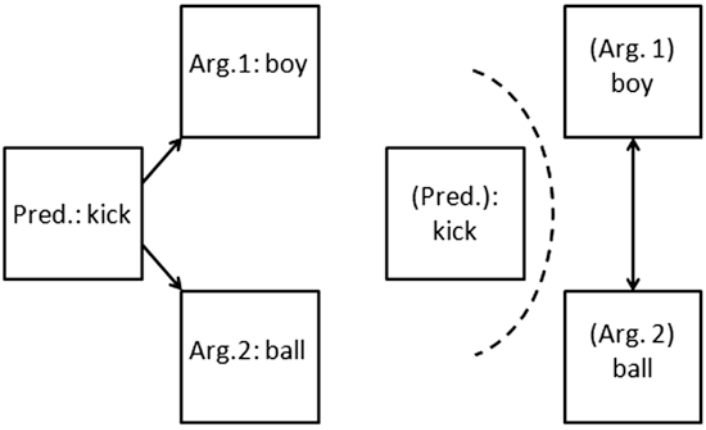


Fig. 7.1 The combinatorial scheme according to the predicate-argument pattern. The scheme on the left represents the indispensability of the predicate according to the concept of argument, while the right one wants to suggest how it is possible to think the two arguments, which strictly speaking would not be such, without involving a third element

In a hypothetical case of the type 'daddy give', where the action (the predicate) is represented by the verb and the argument by noun or gesture, how does one establish whether the argument plays the role of agent or dative (the category X + dative-Schlesinger, 1971-)? How does one know if, respectively, it is /daddy give/ or /give to daddy/? Is this not by chance the most important survey when one intends to go from bimodality to unimodal expression? The recognition of predicate-arguments does not allow us to see what role the argument has in each case: whether it is the grammatical subject or the indirect object. Nor does it allow us to see which types or tokens of argument are expressed with gestures or with verbalization, whether this happens regularly or accidentally, whether the same and only argument expressed with a noun is complemented by a determinative sign or a second noun (and a long etcetera).

To summarize, the use of the Özçalışkan and Goldin-Meadow scheme is legitimate, but (i) it hastily injects certain assumptions into the combinations, suggesting the child somehow distinguishes between two logical types of content, and (ii) it fails to capture the semantic richness present. *Real* semantics is buried under a layout that does not address the relations and dissociations that the child has been able to manage (links of the agent-object-dative or similar type) and that only retains the movement between only two general formal categories. The semantic role model, on the other hand, moves on the surface of what is actually communicated.

Recreating the communicative experience of infants using predicates and arguments forces us to assume them capable of an internalist logical-grammatical type of mental ordering. A more ecological version suggests that, on the contrary, the child relies his communicative efforts on 'external', physical, extra-mental, simply perceived entities, leaving the work involved in understanding to adults, to whom he/she is bound by the ability to jointly attend to a referent. Corporeality, direct

contact with objects and the understanding of intentionality are enough for the child to use, in isolation or in combination, a minimum of signs, which he addresses to the interlocutor with the confidence that, in decoding, the latter will know how to use each element of the context. This ecological framing of development is possibly a preferable alternative to the 'hard' cognitivist mental mapping of predicates and arguments. Rather than an internal, mentalistic recreation that structures *who-does-what-action-on-what-thing-and-for-whom* or *by-which-means* (among other possible complements) through symbols or pure cognitive contents (a sort of map or abstract representation of the situation), we can think of a conjunction of effective signs and circumstantial implicatures that operate in an integrated manner. The expression as such is only one more component of the total scene, moving in its environment, having at its side everything it needs for a successful decoding.

The entire predicate-argument approach is oriented toward the sentential construction. Literally:

> The children produced a small number of adjective + argument combinations ("big gum", "all-sticky" [point to marker]) which were excluded from the analyses; if these combinations are included as predicate + argument constructions, the patterns described below do not change. In addition, combinations containing fillers (e.g., filler + argument: "Sorry Zoe", "please + point at cookie, or filler + predicate: "Please help", "please" + give gesture) were excluded from the analyses because they do not constitute sentential constructions (although they too appeared in gesture + speech before speech alone) (Özçalışkan & Goldin-Meadow, 2005 B105).

If the ontogenesis of communication is to be studied, focusing on the roles played by each of the effectively produced signs in the different expressions will prove more fruitful than attending to the *plausible* logical-grammatical organization of a combination. Coding by semantic roles (a proposal we adopt and extend to gesture-word combinations) is based on the understanding of what the child sees occurring in his immediate environment and conceives signs as components of action schemes and intersubjectively organized routines. Many years ago, one of us (Español, 2004) explored the possibility of using the case grammar scheme devised by Fillmore to interpret the first substitutions made by children in communicative contexts. She verified that after performing *recognition gestures* (gestures repeating the morphology of familiar actions for no practical purpose: bringing an empty spoon to the mouth), children replace the objects of this recognition by other functionally inadequate ones, obviating their physical properties or their correspondence with the action in question. Gestures with changed objects were baptized *permuted gestures*. By using them, children make the link between the sign and its referent and take a step in the direction of fictional play. The permutation usually concerns the objects of a given action (e.g., a spoon is used for the action of combing). At the age of 1; 10 (19), a girl picks up from the floor some strands of hair that have just come off her doll's head, places them on a spoon and offers them to that same doll as a food. According to Español, each element of the scene played a semantic role: "In terms of the grammar of cases (...): The Action: is to feed. The agent of the Action: is the child. The instrument: is the spoon. The recipient of the Action: is the doll. The object: is the hair" (Español, 2004, p. 243). Case grammar thus revealed its aptitude

to explain the first symbolic actions without the need for grammatical speech. This incipient brush with Fillmore's categories made it possible to understand his cases as a derivative of the way in which young children naturally replace the elements of everyday routines with others.

Substitutions of one type or the other are enhanced when the expression involves two signs that unite in a combination. If we consider that in children's productions prior to word couplings, gesture-word constructions are present, and that gestures, for example, deictic gestures, play the role of agent, object, dative, and so on, investigating how gestures and orality *intermodally* distribute the roles is mandatory.

Fillmore presents a difficulty because its categories are derived from what adult speech grammar dictates. The benefit of classifying bimodal utterances by semantic roles of combined signs (thematic duples, triples, tetrads, etc.) arises from the fact that the categories created by Brown were drawn from a casuistry of actual verbal uses. However, this advantage should not disguise the fact that these categories are also forged for grammar (in this respect they coincide with the predicate-argument model). Due to this linguistic orientation, they imply certain proscriptions that must be overcome in the interest of research on the ontogenetic development of communication, incorporating gestures as well as other types of expression: facial, bodily, etc. Because of this linguistic orientation, they imply certain proscriptions that must be overcome in the interest of research on the ontogenetic development of communication, incorporating not only gestures but also other types of expression: facial, corporal, etc. The inclusion of gestures and other signs could result in the duples that in Brown had gathered 70% of the children's verbalizations being affected in one way or another. By including other variables, it could happen that the word-word combinations of low frequency in Brown would suddenly gain more prominence, or that it would be necessary to add new duples. An expression that includes a vocative (and that could only be avoided by thinking exclusively of propositional utterances) implies in itself a new signic duple (or more than one, depending on whether it is necessary to establish different pairs: vocative + agent, vocative + locative, etc.). The combination "Like this" + X (as in observation 3) is not covered by Brown either, nor are the bimodal cases of molecular proto-proposition (as suggested for observations 12 and 15–17).[9] Combinations formed with vocalizations or interjections are also missing. With a more generous scope of signs and combinations we can speculate on an unprecedented fruitfulness for gesture-word relations research.

To obtain an atlas containing all variants of communication, the categories of Brown must not only be pulled out of their unimodal confinement and exploited for bimodal combinations, but also extended to all dimensions of semiotic interaction, for they all cooperate, intertwine and underpin each other. This could prevent communicative skills research from ending with supplementary combination (of two or

[9] The combinations of two simultaneous signs with antithetical senses could never become visible in the strictly linear oral modality, and would never have been noticed from Brown's point of view.

more components) and the research on language sources from directly shifting from vocalizations to sentential constructions directly.[10]

Pending such a comprehensive system of observation categories, Table 7.1 comparatively presents the categories created by Brown (unimodal) with examples, the predicate and argument categories (bimodal), with examples, and again the categories of Brown as a reference for a bimodal approach, also with examples, in this instance, coming from our particular *corpus*.

7.5 Toward an Ecological Code of Developmental Communication

Given the limitations still offered by Brown's categorization in the interest of investigating pregrammatical communications, the evaluation of bimodal combinations by semantic roles should only be taken as a transitional tool. In communication phenomena, not only gestural or oral signs, whose communicative character is beyond discussion, but also other behavioral phenomena that, in certain circumstances, are cooperatively associated with the communicative act, usually appear as elements in a plot. As if, when produced, communication rests on certain behavioral components, surrounded by conditioning factors that are not per se intentional signs, but which at the same time could not be dissociated from the fact of actual communication without putting it at risk. It is not only that any event, thing, or punctual aspect of the world can become a sign and address an interlocutor. It is also not about the background that the sign needs to be contextualized and to have an occasional meaning or, according to Peirce, a *dynamic interpretant*. Beyond all this, some behavioral elements lend signs a support or underpinning by which a communicative network or mesh is forged between them, between communicative signs themselves and other auxiliary materials. As Kendon highlighted: "However, at the same time, action in other parts of the body concurrent with the production of signs may also be observed. This action, though most of it is not part of the sign lexicon, is nevertheless taken into account by the recipient of the utterance, and plays an important part in the way the total meaning of the utterance is conveyed" (Kendon, 2020, p. 75). In this way, body expressiveness, for example (which we previously disassociated from gestural signs), can be added to a certain communication and, so to speak, *totalize* it, become part of the *semantic whole* elaborated for a semiotic partner.

[10] Body and facial expressive manifestations accompanying vocalizations of various kinds are not considered as forming combinations. As indicated (Sect. 7.2), for the former, demarcating their "good form", their beginning and end, their meaningful content is not always possible, as they tend to interweave with each other in a hardly analyzable continuity. For the latter, the problem is that they are usually reflex responses and therefore go beyond the horizon of intentional behavior. However, within the appropriate context, they can both be part of communicative expressions and can decisively contribute to their meaning.

Table 7.1 Brown's categories and bimodal expressions

Brown's categories	Brown's examples[a]	Özçalışkan & Goldin-Meadow's categories	Özçalışkan & Goldin-Meadow's examples[b]	Bimodal examples[d]
Nomination	This; this + (noun)	(not applicable or up to circumstance)	Car (point to car)[c]	This + pointing to object
Recurrence	More + (noun, verb)	(not applicable or up to circumstance)	–	Mom + reaching gesture (towards the object)
Nonexistence	No more; dog away	(not applicable or up to circumstance)	–	_oodle /noodle/+ negation through head movement
Agent-action	Car go; mommy fix (205)	Predicate + argument	Baby + EAT (iconic)	Me + falling backwards gesture
Action-object	Hit ball	Predicate + argument	Eat muffin (hold-up)	Reaching gesture + name of the object
Agent-object	Mommy pumpkin	Argument-argument	Mommy + CUP (point)	Point to dad + car (asking)
Action-locative	Put floor	Predicate + argument	Daddy gone + outside (point)	Marco, am! + pointing to location (kitchen)
Demonstrative-entity	This car	Argument-argument[c]	–	This -ookie /cookie/
Entity-locative	Baby table	Argument-argument	a choochoo train' + TUNNEL (point)	Hand to chest + pointing a location/ me there/
Entity-attribute	Dirty my hands (p. 209)	Argument-argument	All-sticky" (point to marker)	Nice" + shows object
Possesor-possesion	Daddy chair	Argument-argument		Me/ mine/ + pointing the object

Exemplification of Brown's categories for bimodal expressions according to Özçalışkan & Goldin-Meadow and our own observations

[a]Examples taken or reconstructed from the formulas by Brown (Brown, 1973, pp. 193–209)

[b]Examples taken from Özçalışkan & Goldin-Meadow, 2005 and Özçalışkan & Goldin-Meadow, 2009. The column is not complete because these articles have different aims than our chapter and do not provide examples for each Brownean duple. We respect here the capitalization of the original

[c]Only in cases where the demonstrative is not redundant with the entity term

[d]Examples taken from Rodríguez, 2021

For what has been said, Brownian categories, whose usefulness we sought to emphasize, will serve us to study the transition from bimodal combinations to morphosyntactic structures in the best possible way; but in order to more fully address the phenomena of communication and its transformations in ontogenetic development, we need to broaden the field of observation to include all factors that, in one way or another, are part of the ability to signify with intention. Evidently, it will be

necessary to transcend the limits of gesture and orality, because interaction simultaneously follows other converging paths. For example:

> Observation 15: 1; 9 (2). Bruno is sitting on the floor, almost lying down, playing with a friction car, dragging it forcefully backwards and letting it go. His father is with him, watching the game behind his back and making sporadic observations. To make the car go farther and farther Bruno moves the car with excessive force, until the movement ends in a breaking sound. The car no longer goes 'brrrmmm' nor can it move forward. Then Bruno says one word: "Go", with a questioning tone and turning to the father (torso rotation). His eyebrows are raised, as if waiting for some kind of solution. The word, prosodically modulated, and the facial expression would have been enough for the message to be understood, but immediately Bruno makes a shaking motion with his head. Then he repeats three times the action of propelling the broken-down car, with three simultaneous onomatopoeias ('brrrmmm!'). Between each of these sets of movement and onomatopoeia he gazes the father no longer with an interrogatory attitude, but as if he were informing him of the concrete facts that he was reproducing with the gesture of his hand.

The reference action, 'go', and the interrogation tone, intertwine with the gaze, which is also, as in the intonation, a way of involving the partner. The idea of /go/ is in turn in the onomatopoeia of the roaring engine, supported by the representational gesture of the movement of the toy before the accident (there is an eagerness to imitate reality in all its facets). Negation – reversing the affirmation of 'go' or transforming the meaning of the question – is informed with a head gesture. In the oral form, two distinct signs follow and replace each other: the allusion to the passing of the car and the sound of the engine. Concerning body participation, we must group together the facial expression – questioning the interlocutor with full awareness –, the torso rotation and, finally, the three hand gestures emulating the regular car ride.[11] The whole message can be graphed as in Fig. 7.2:

The tendency to take the verbal expression as the strategic center of a meaningful combo could suggest that the word 'go' fulfills this functionality, but, under a multimodal analysis, it is seen that the center is, on the contrary, the idea of /not go/ (the

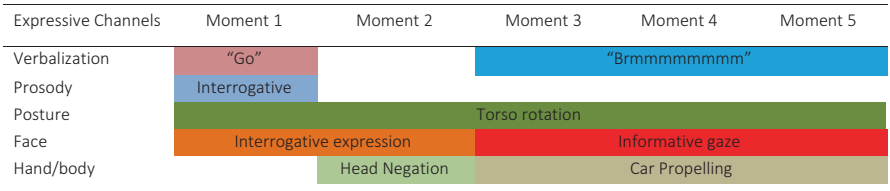


Fig. 7.2 Observation 15 analysis. The horizontal lines indicate the different expressive channels involved in this communicative episode. The colored stripes indicate the duration of each of the expressive units used in five different moments (columns) of the total message structure

[11] The body movements contemplated here should not be confused with the expressive movements (or body expression) discussed in Sect. 7.1 and which have been left out of the gestures. The positioning of the torso in observation 15 is not an expressive type of movement, but it is a movement that, in context, is coupled with other elements and composes the communicative act, making it clear to whom the specific message of verbalization and gestures is addressed. Without elements of meaning, the movement as such would be only part of the behavior, but attached to them it is integrated cooperatively to the communicative act.

negation, as principal as the action denied, is expressed by means of a gesture). The object of the case, the car, appears later and by two different means: the representational gesture of movement and the sound of the engine (a complementary combination). At the same time, and in conjunction with the first sign emitted, the word, Bruno uses the tone of voice to bring a pragmatic modulation to the message. The intonation matches, on its side, the motif present in the facial expression. The intonation matches, on its side, the motif present in the facial expression.

From the very beginning, the content of the message lies in the intention of the sender, disaggregated in intersecting solidary channels materializing the underlying idea. The entire signification unfolds in different concurrent signs, which induces us to think of communication not as a linear progression but rather as a *pentagram.* In it we find, taking into account each parallel line, at each instant, so to speak in each compass, the harmony among the different concerted elements, the imaginary threads that vertically link what is happening in each diachronic or horizontal axis. A pentagram of this condition (or a hexagram, a heptagram, etc., depending on the total number of communicative lines to be analyzed) not only challenges the aforetime linearity Saussure assumed for language (Rodríguez, 2018b), but also reflects with the utmost justice the semiotic facticity as we verified it in its early reality. Far from being a purely denominational witticism, the notion of pentagram alludes, of course, to the foundations of musicality that lie in the most primordial baby-adult exchanges, and at the same time rescues the idea of the composition of all the resources at the service of the communicative message. The *multimodal weft* of different integrated expressive lines is not only a passing support for a verbality yet to be emancipated from gestures and other forms of signifying, but the processing structure upon which all other semantics can eventually be established. Like in a musical pentagram, what emerges at each instant from a single communicative act cannot be explained as an aggregate but from its multi-axial nature.

A more complete description of infant communicative development should tend toward a category system capable of containing most of the varieties of signic combinations. As we have indicated, semantic roles in the Brown style may take us beyond the bimodal models used to date (supplementary combinations, predicate-argument patterns), but encompassing communication in all its plurality of aspects and component factors requires a model that includes vocalizations, body language, suprasegmental phonemes, onomatopoeias and postural attitudes, as well as any other semiotic possibility used by senders toward interlocutors (contemplating the concomitant phenomena which, although they may not be in themselves communicative, are indissolubly intertwined with communicative signs and compose, in one unit, an irreducible *semiotic quantum*). If the child's communication in gestures and words leads as a rule to grammar, it should not be forgotten that it should be thought of at all times not as grammar, which is necessarily unimodal, but rather as multimodal *penta-grammar*.

7.6 Final Remarks

In this chapter, we have discussed the notion of gesture and the different ways it has been understood in developmental psychology. We then explained why we prefer to think of it as an emergent of prior processes and skills, essentially linked to a communicative intention, and how, as such, it is integrated into pregrammatical bimodal combinations. Within the bimodal combinations, we have surveyed marginal cases and discussed the benefit of including couplings with vocative signs, negation (in all its variants) and expressive signs in forthcoming studies. In fact, negation allowed us to propose that in the simultaneity of gesture and word we already find expressions preluding molecular logical-linguistic constructions. Likewise, in this analysis we have isolated the double gesture from the compound gesture and exemplified a curious cross-modal conversion between a verbal input that a child then expresses as a gesture (observations 12 and 13), supporting, in a novel way, the idea of an originally bimodal sign processing mechanism (as postulated by Kendon and McNeill).

Our analysis of bimodal combinations led us to review the so-called supplementary combinations and to recover Brown's semantic-pragmatic pairs, as a more flexible and appropriate tool for tracking progress from gesture-word combinations to the threshold of grammatical articulation. This system of categories makes it possible, we argue, to take the child's expression at the level where his experiences are semiotically elaborated in a primary form, that is, at the level where perceptual and agentive experiences are organized as signs for an interlocutor. The Brown-style categories allow us to register how gesture and word, and then only words, divide the message into different sign contents. Thus, they allow us to explore how the cognitive-communicative system *assigns* (in the most appropriate and literal way) a gestural or verbal surrogate element to each content. It follows that role semantics is better suited than other systems to obtain a reliable record of their distribution within the message, and would thus better enable a detailed tracing of ontogenetic evolution in the field of communication.

Throughout the transition between bimodal, then unimodal, and finally grammatically regulated sign combination communications, role mapping underlies these three alternatives and, as Fillmore was able to point out, governs even in the terminal stage. If this is so, it is because through words and gestures and under morphosyntactic structures there is invariably a certain situational-pragmatic understanding of the experiences in which children actively learn to engage.

This vast research territory stems from restoring gesture as a co-protagonist of human communication in natural exchange situations, with body-present interlocutors, and particularly in the context of native language acquisition. But gesture has also been highlighted as a comparative feature with nonhuman primate species. We can see here that its role as a communicative agent does not depend on language or the logical form of the proposition. In the same vein, we can recall an anecdote where the economist Pietro Sraffa, member of the Cantabrigian Cafeteria group, would have forced Wittgenstein to reconsider his entire Tractarian philosophy. As

the story goes (see Malcolm, 2001; McGuinness, 2008), during a conversation on the grammar of propositions and its correlation with world facts, Sraffa gave the philosopher a Neapolitan gesture of disdain or displeasure, with his fingertips running below his chin from the inside out, and then asked him what would be the grammar involved in that. Perhaps not everything was grammatical, as the Viennese had assumed. We can see how, in many different contexts, gesture has provided a way out of overly logicist, fallaciously descriptive versions of language, and how, particularly from early childhood exploration, has reshaped the phenomenology of semiosis in human subjects.

It remains a pending challenge to widen the margins of bimodal combinations. With this aim in mind, we wanted to emphasize, lastly, that if grammar corresponds to a semiotic-expressive emancipation from previous competences (emancipation and, of course, reorganization), it would benefit from a research that opens the diaphragm of observation to the maximum and considers all kinds of signs. Developmental psychology has long emphasized the importance of early musicality and social-cognitive skills in understanding the general cognitive processes of infants and early children. According to this approach, studies on children's communicative skills will be enriched by an approach that is not limited to combinations and can think of that basic, multimodal, *pentagrammatic* harmony that lies at the genesis of semiosis.

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and linguistics compass, 9*(11), 437–451. https://doi.org/10.1111/lnc3.12168

Andrén, M. (2014). On the lower limit of gesture. In M. Seyfeddinipur & M. Gullberg (Eds.), *From gesture in conversation to visible actin as utterance. Essays in honor of Adam Kendon* (pp. 153–174). John Benjamins Publishing Company.

Bates, E. (1976). *Language and context: The acquisition of pragmatics*. Academic Press Inc.

Bates, E. (1979). *The emergence of symbols*. Academic Press.

Birdwhistell, R. L. (1970). *Kinesics and context: Essays in body motion communication*. University of Pennsylvania Press.

Bloom, L. (1970). *Language development: Form and function in emerging grammars*. MIT Press.

Bowerman, M. (1973). *Early syntactic development: A cross linguistic study with special reference to Finnish*. Cambridge University Press.

Brown, R. (1973). *A first language. The early stages*. Harvard University Press.

Bullowa, M. (Ed.). (1979). *Before speech. The beginning of interpersonal communication*. Cambridge University Press.

Butcher, C., & Goldin-Meadow, S. (1993). *From one spoken word to two: exploring the changing nature of gesture*. S/D. https://files.eric.ed.gov/fulltext/ED356898.pdf

Capirci, O., & Volterra, V. (2008). Gesture and speech: The emergence and development of a strong and changing partnership. *Gesture, 8*(1), 22–44. https://doi.org/10.1075/gest.8.1.04cap

Capirci, O., Iverson, J. M., Pizzuto, E., & Volterra, V. (1996). Gestures and words during the transition to two-word speech. *Journal of Child Language, 23*(3), 645–673. https://doi.org/10.1017/S0305000900008989

Capirci, O., Contaldo, A., Caselli, M. C., & Volterra, V. (2005). From action to language through gesture: A longitudinal perspective. *Gesture, 5*(1–2), 155–177. https://doi.org/10.1075/gest.5.1.12cap

Capobianco, M., Pizzuto, E. A., & Devescovi, A. (2017). Gesture-speech combinations and early verbal abilities: New longitudinal data during the second year of age. *Interaction Studies, 18*(1), 55–76. https://doi.org/10.1075/is.18.1.03cap

Cartmill, E. A., Beilock, S., & Goldin-Meadow, S. (2012). A word in the hand: Action, gesture and mental representation in human and non-human primates. In J. Steele, P.F. Ferrari, & L. Fogassi (Eds.), *From action to language: Comparative perspectives on primate tool use, gesture and the evolution of human language* (pp. 129–143). *Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585). https://doi.org/10.1098/rstb.2011.0162.

Clark, R. A. (1978). The transition from to action to gesture. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 231–257). Academic Press.

Cooperrider, K., Abner, N., & Goldin-Meadow, S. (2018). The palm-up puzzle: Meanings and origins of a widespread form in gesture and sign. *Frontiers in Communication, 3*, 23. https://doi.org/10.3389/fcomm.2018.00023

Corballis, M. (2002). *From hand to mouth. The origins of language*. Princeton University Press.

Efron, D. (1941/1972). *Gesture, race and culture*. Mouton and Co.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49–98.

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana*. Antonio Machado.

Fillmore, C. (1968). The case for case. In E. W. Bach & R. T. Harms (Eds.), *Universals in linguistic theory* (pp. 1–88). Holt, Rinehart, and Winston.

Gardner, R. A., & Gardner, B. T. (1969). Teaching sign language to a chimpanzee. *Science, 165*, 664–672.

Goldin-Meadow, S., & Brentari, D. (2017). Gesture, sign, and language: The coming of age of sign language and gesture studies. *Behavioral and Brain Sciences, 40*, 1–17. https://doi.org/10.1017/S0140525X15001247

Hall, E. T., & Trager, G. L. (1953). *The analysis of culture*. American Council of Learned Societies.

Hewes, G. (1973). Primate communication and the gestural origins of language. *Current Anthropology, 14*, 5–24.

Iverson, J., Capirci, O., & Caselli, M. (1994). From communication to language in two modalities. *Cognitive Development, 9*, 23–43. https://doi.org/10.1016/0885-2014(94)90018-3

Karmiloff-Smith, A. (1992). *Beyond modularity: A developmental perspective on cognitive science*. MIT Press.

Karousou, A. (2003). *Análisis de las vocalizaciones tempranas: su patrón evolutivo y su función determinante en la emergencia de la palabra [Analysis of early vocalizations: their evolutionary pattern and their determinant role in the emergence of speech]*. (Doctoral tesis). Complutense University of Madrid.

Karousou, A., & López-Ornat, S. (2013). Prespeech vocalizations and the emergence of speech: A study of 1005 Spanish children. *The Spanish Journal of Psychology, 16*, 1–21. https://doi.org/10.1017/sjp.2013.27

Kelly, S. D., Iverson, J. M., Terranova, J., Niego, J., Hopkins, M., & Goldsmith, L. (2002). Putting language back in the body: Speech and gesture on three time frames. *Developmental Neuropsychology, 22*(1), 323–349. https://doi.org/10.1207/S15326942dn2201_1

Kendon, A. (1972). Some relationships between body motion and speech. An analysis of an example. In A. Aiegman & B. Pope (Eds.), *Studies in dyadic communication* (pp. 177–210). Pergamon Press.

Kendon, A. (1980). Gesticulation and speech: Two aspects of the process of utterance. In M. R. Key (Ed.), *The relationship of verbal and nonverbal communication* (pp. 207–227). Mouton and Co.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kendon, A. (2020). *Sign language in Papua New Guinea. A primary sign language from Upper Lagaip Valley, Enga Province*. John Benjamins Publishing Company.

Kersken, V., Gómez, J. C., Liszkowski, U., Soldati, A., & Hobaiter, C. (2019). A gestural repertoire of 1- to 2-year-old human children: In search of the ape gestures. *Animal Cognition, 22*, 577–595. https://doi.org/10.1007/s10071-018-1213-z

Kita, S. (Ed.). (2003). *Pointing. Where language, culture, and cognition meet*. Erlbaum.

Kovács, A. M., Tauzin, T., Téglás, E., Gergely, G., & Csibra, G. (2014). Pointing as epistemic request: 12-month-olds point to receive new information. *Infancy, 19*(6), 543–557. https://doi.org/10.1111/infa.12060

Liebal, K., & Call, J. (2012). The origins of non-human primate's manual gestures. In J. Steele, P.F. Ferrari, & L. Fogassi (Eds.), *From action to language: Comparative perspectives on primate tool use, gesture and the evolution of human language. Philosophical Transactions of the Royal Society B: Biological Sciences*, *367*(1585), 118–128.. https://doi.org/10.1098/rstb.2011.0044.

Liszkowski, U., Carpenter, M., Henning, A., Striano, T., & Tomasello, M. (2004). Twelve-month-olds point to share attention and interest. *Developmental Science, 7*(3), 297–307. https://doi.org/10.1111/j.1467-7687.2004.00349.x

Lock, A. (1978). The emergence of language. In A. Lock (Ed.), *Action, gesture and symbol: The emergence of language* (pp. 3–18). Academic Press.

Lüke, C., Ritterfeld, U., Grimminger, A., Rohlfing, K. J., & Liszkowski, U. (2020). Integrated communication system: Gesture and language acquisition in typically developing children and children with LD and DLD. *Frontiers in Psychology, 11*, 118. https://doi.org/10.3389/fpsyg.2020.00118

Malcolm, N. (2001). A memoir. In N. Malcolm (Ed.), *Ludwing Wittgenstein. A Memoir* (pp. 21–84). Oxford University Press.

McGuinness, B. (2008). *Wittgenstein in Cambridge: Letters and Documents 1911–1951*. Blackwell.

McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. The University of Chicago Press.

McNeill, D. (2005). *Gesture and thought*. University of Chicago Press.

McNeill, D. (2007). Gesture and thought. In A. Esposito, M. Bratanic, E. Keller, & M. Marinaro (Eds.), *Fundamentals of verbal and nonverbal communication and the biometric issue* (pp. 20–33). IOS Press.

McNeill, D. (2012). *How language began: Gesture and speech in human evolution*. Cambridge University Press.

McNeill, D. (2017). Gesture-speech unity: Where it is, where it came from? In R. Church, M. Alibali, & S. Kelly (Eds.), *Why gesture: How the hands function in speaking, thinking and communicating* (pp. 77–101). John Benjamin Publishing Company.

Moreno-Núñez, A., Murillo, E., Casla, M., & Rujas, I. (2021). The multimodality of infant's rhythmic movements as a modulator of the interaction with their caregivers. *Infant Behavior and Development, 65*(101645), 1–8. https://doi.org/10.1016/j.infbeh.2021.101645

Müller, C. (2018). Gesture and sign: Cataclysmic break or dynamic relations? *Frontiers in Psychology, 9*, 1–20. https://doi.org/10.3389/fpsyg.2018.01651

Murillo, E., & Belinchón, M. (2013). Patrones comunicativos multimodales en la transición a las primeras palabras: cambios en la coordinación de gestos y vocalizaciones [Multimodal communicative patterns in the transition to first words: changes in the coordination of gestures and vocalizations]. *Infancia y Aprendizaje, 36*(4), 473–487. https://doi.org/10.1174/021037013808200258

Murillo, E., Montero, I., & Casla, M. (2021). On the multimodal path to language: The relationship between rhythmic movements and deictic gestures at the end of the first year. *Frontiers in Psychology, 12*, 1–8. https://doi.org/10.3389/fpsyg.2021.616812

Özçalışkan, Ş., & Goldin-Meadow, S. (2005). Gesture is at the cutting edge of early language development. *Cognition, 96*(3), B101–B113. https://doi.org/10.1016/j.cognition.2005.01.001

Özçalışkan, Ş., & Goldin-Meadow, S. (2009). When gesture-speech combinations do and do not index linguistic change. *Language and Cognitive Processes, 24*(2), 190–217. https://doi.org/10.1080/01690960801956911

Özçalışkan, Ş., & Goldin-Meadow, S. (2010). Sex differences in language first appear in gesture. *Developmental Science, 13*(5), 752–760. https://doi.org/10.1111/j.1467-7687.2009.00933.x

Pollick, A. S., & De Waal, F. B. (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Sciences, 104*(19), 8184–8189. https://doi.org/10.1073/pnas.0702624104

Pouw, W., Harrison, S. J., Esteve-Gibert, N., & Dixon, J. A. (2020). Energy flows in gesture-speech physics: The respiratory-vocal system and its coupling with hand gestures. *The Journal of the Acoustical Society of America, 148*(3), 1231–1247. https://doi.org/10.1121/10.0001730

Rodríguez, F. G. (2018a). La pragmática de la comunicación antes del lenguaje [The pragmatics of communication prior to language]. *Revista IRICE, 35*, 95–128.

Rodríguez, F. G. (2018b). El pentagrama multimodal y la unidimensionalidad del lenguaje. El aporte de la psicología del desarrollo a la más plena comprensión de la comunicación verbal [The multimodal pentagram and the unidimensionality of language. The contribution of developmental psychology to a deeper understanding of verbal communication]. *Lengua y habla, 22*, 203–225.

Rodríguez, F. G. (2021). *Semántica y pragmática en la comunicación del niño pregramatical. Gesto y palabra antes de la articulación morfosintáctica* [Semantics and pragmatics in the communication of the pregrammatical child. Gesture and speech before morphosyntactic articulation]. Teseo.

Rodríguez, F. G., & Español, S. (2016). Composiciones bimodales de gesto y vocalización en el inicio de la comunicación verbal [Bimodal compositions of gesture and vocalization at the beginning of verbal]. *Infancia y Aprendizaje, 39*(4), 661–693. https://doi.org/10.1080/02103702.2016.1215086

Rodríguez, F. G., & Español, S. (2019). Communicative functions during the pre-grammatical stage: a case study/Las funciones comunicativas durante la etapa pregramatical: un estudio de caso. *Infancia y Aprendizaje, 42*(2), 413–458.

Rosch, E. (1973). Natural categories. *Cognitive Psychology, 4*, 328–350.

Sapir, E. (1927/1951). The unconscious patterning of behavior in society. In D. G. Madelbaum (Ed.), *Selected writings of Edward Sapir in language, culture and personality* (pp. 544–559). University of California Press.

Sarriá, E., & Rivière, Á. (1991). Desarrollo cognitivo y comunicación intencional preverbal: un estudio longitudinal multivariado. *Estudios de Psicología, 46*, 35–52.

Sarriá Sánchez, E. (1991). Observación de la comunicación intencional preverbal: un sistema de codificación basado en el concepto de categoría natural. *Psicothema, 3*(2), 359–380.

Schlesinger, I. M. (1971). Production of utterances and language acquisition. In D. Slobin (Ed.), *The ontogenesis of grammar* (pp. 63–101). Academic Press.

Stern, D. (2010). Forms of vitality. In *Exploring dynamic experience in psychology, the arts, psychotherapy and development*. Oxford University Press.

Stokoe, W. C. (1960). *Sign language structure*. Linstock Press.

Trevarthen, C., Delafield-Butt, J. T., & Schögler, B. (2011). Psychobiology of musical gesture: Innate rhythm, harmony and melody in movements of narration. In A. Gritten & E. King (Eds.), *Music and gesture II* (pp. 11–43). Ashgate.

Vihman, M., DePaolis, R. A., & Keren-Portnoy, T. (2009). A dynamic systems approach to babbling and words. In E. L. Bavin (Ed.), *The Cambridge handbook of child language* (pp. 163–184). Cambridge University Press.

Vilà-Giménez, I., & Prieto, P. (2021). The value of non-referential gestures: A systematic review of their cognitive and linguistic effects in children's language development. *Children, 8*(2), 148. https://doi.org/10.3390/children8020148

Volterra, V., Caselli, M. C., Capirci, O., & Pizzuto, E. (2005). Gesture and the emergence and development of language. In M. Tomasello & D. Slobin (Eds.), *Beyond nature-nurture: Essays in honor of Elizabeth Bates* (pp. 3–40). Lawrence Erlbaum.

Volterra, V., Capirci, O., Caselli, M. C., Rinaldi, P., & Sparaci, L. (2017). Developmental evidence for continuity from action to gesture to sign/word. *Language, Interaction and Acquisition, 8*(1), 13–41. https://doi.org/10.1075/lia.8.1.02vol

Werwach, A., Mürbe, D., Schaadt, G., & Männel, C. (2021). Infants' vocalizations at 6 months predict their productive vocabulary at one year. *Infant Behavior and Development, 64*, 101588. https://reader.elsevier.com/reader/sd/pii/S0163638321000631?token=934521CE33EF130F8EF0443A4D40952F4EC36F593026600B6AEF051C338F86755EE010F9F28A7E740487BE6F27A69046&originRegion=us-east-1&originCreation=20211201212428

Fernando G. Rodríguez is PhD in Educational Sciences from Universidad de Buenos Aires. He has carried out research and training stays in Germany (DAAD) in Linguistics and Semiotics. He is currently Professor of Semiology and Human Communication. He is member of the Research Group of Socio-Cognitive Processes in Infancy and Early Childhood at the Psychology of Knowledge and Learning Department (FLACSO, Argentina). As a researcher, his work is concerned with the development of communicative skills in young children, the structure of semiotic processes in general and the relationships between mind, sign and corporeality. Author of *Semántica y pragmática en la comunicación del niño pregramatical. Gesto y palabra antes de la articulación morfosintáctica* [Semantics and pragmatics in the communication of the pregrammatical child. Gesture and speech before morphosyntactic articulation] (2021) and co-editor of *Los signos del cuerpo. Enfoques multimodales de la mente y el lenguaje* [The signs of the body. Multimodal approaches to mind and language] (2018). His publications can be found at https://www.aacademica.org/fernando.gabriel.rodriguez

Silvia Español has a PhD in Psychology from Universidad Autónoma de Madrid (Spain). She is a researcher at the Latin America Institute of Social Research (IICSAL), Faculty of Latin America Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET, Argentina). She is a certified Somatic Educator in the Feldenkrais Method. She coordinates the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc in Cognitive Psychology, and Director of FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. Her work deals with cognitive developmental psychology, psychology of music, and the study of human movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/silvia.espanol

Chapter 8
Multimodal Expression in Communicative Functions, Gestures, Vocalizations, and the Contribution of Early Musicality

Fernando G. Rodríguez

Abstract The communicative functions with which infants and pregrammatical speaking children use signs to interact with others have been the subject of different classifications, usually concentrated in one modality (gesture, word) or in bimodal combinations of gesture and word, or oriented by a particular criterion (for example, some were conceived thinking of types of signs, others of the illocutionary acts of utterances, etc.). With an integrative aim, we propose a classification of the communicative functions of oral, gestural, and combined modalities, also suitable to include signs of other types and especially designed to deal with pregrammatical expressive resources. The system of categories is forged in attention both to the structural scheme of the communicative act, bearing in mind its actors, roles, and component elements, and to the motives or intentions that lead pregrammatical infants and children to interact with others through behaviors performed, purposively, for the other to decode. This double consideration, together with the longitudinal recording of a single case during the period of first word combinations, allowed both to include categories usually omitted in classifications for competent speakers (adults, older children) and to exclude others that still require the acquisition of prior enabling cognitive skills. Three new categories are also posited, two of them liminal to the communication phenomenon, linked to musicality and multimodality, which allow conceiving the development of the child's semiosis as a continuity between precommunicative and later communicative dyadic behavioral skills.

Keywords Communicative functions · Precommunicative functions · Expression · Musicality

F. G. Rodríguez (✉)
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_8

8.1 Introduction

The idea of communication in early childhood is divided between two conceptions: that which implies connection or interaction between adult and child, a mutual antepredicative understanding, and that which implies communication in a restricted sense, defined by the trait of deliberately exchanging signs about something (having intention, or *intentio*, in the meaning rehabilitated by Brentano in the second half of the nineteenth century). Put in other terms, there is, on the one hand, communication between individuals, where specifically saying 'between' excludes all content, and there is, on the other hand, thematic communication, deployed of course within the space *between-two*, but where there is room for a *what* or *aboutness*. If the first six chapters of the book that the reader has in his hands are devoted to the first version of communication, and if in Chap. 7 we started to talk about communication in a different sense (thematic communication), this chapter will continue with this second acceptation.

Under this perception, the operations by which the baby manages to understand that a fact refers to such-and-such other fact−that a movement, a facial expression, a particular morphology adopted by the hand, a glance, a turn of the head refer to something they are not−are essentially linked to social and learning situations. If an adult makes a pointing gesture at the baby's sight, he/she will not achieve anything other than the attention on the tip of his/her finger. For him to be able to follow the imaginary line between the hand and some referent, it is necessary to have developed enabling skills that can only be explained from previous experience and the social framework. The connections between an expression and a content or empirical correlate start from experience, and the link obtains the necessary reinforcements from physical circumstances, from the agents responsible for the baby's socialization and from cultural determinants (which are tacitly revealed through the action of these same agents). The communicative sign, from its emergence in adult-baby interactions, bears the mark of its fundamental utility, the mutual understanding with the interlocutors (who, literally speaking, only become so through the mediation of sign).

The functions historically distinguished in the study of language, centered on the descriptive-gnoseological uses of the word and the proposition, were later equated with other functions that scholars had considered to play a secondary role (even though Aristotle had already distinguished between a *lógos semantikós* and a *logos apophantikós*, respectively language understood in a broad sense and language only linked to the expression of propositional contents). At the core of the linguistic turn, Wittgenstein had pointed out that language has multiple performances: ordering, recounting an event, inventing a story, making a joke, pleading, thanking, greeting, etc. (Wittgenstein, 1953, §23). In the 1930s, Karl Bühler had suggested an outline ordering for all possibilities of language, whose general functions could be divided around the three grammatical persons. The expressive, appellative, and informative functions would respectively have their center in the first, second, and third persons (Bühler, 1934/1979). Subsequently, linguistics considered it appropriate to

distinguish a metalinguistic function from the informative function, since although both functions are concerned with informing, the latter informs exclusively about the language, which is both the content and the vehicle of expression. Later a poetic function was recognized (Jakobson, 1960), which could be understood as an extension of the expressive element in the subject, a subset within the expressive function, but between these two functions the link is not as immediate as that noted above between the informative and the metalinguistic function. Although it is clear that there can be no poetic function without expressive function, because the latter is the first communicative form of subjectivity, poetic skill requires much more than expressiveness (or else the technical expertise of poetic creation would be within everyone's grasp). In addition, two other varieties of language use constitute the most recognized cast of linguistic functions: the phatic exchange, for which the content of the message is thematically inconsequential, and the performative utterance, which enacts through signs what is uttered (Austin, 1962).

The growing interest in communication outside the limits of verbal expression brought with it, of course, research on the uses of gestures and other non-linguistic signs. In the 1970s, studies on infancy and early childhood identified, based on Austin's three types of speech acts, the successive arrangement of a perlocutionary, an illocutionary, and finally a locutionary phase (Bates et al., 1975, Chap. 2 in this volume). This implied that if in the first instance meaningful acts were simply by-products of infant behavior, they later became intentional and deliberately sought to be interpreted, and ultimately sought to be interpreted through institutionalized conventions that fixed the meaning of each term of expression.

If studies on infancy and early childhood made it possible to convert linguistic functions into communicative ones by revealing the utilities of the baby's expressions at a time prior to the appearance of the word and the proposition, the recognition and examination of bi- and multimodal combinations (to which Chap. 7 in this volume was devoted) made it possible to investigate in greater depth what the pregrammatical child could achieve, in terms of communication, before the juxtaposition of signs was ordered by rules or principles. However, the Brownean or similar categories, however much they may be detached from the propositional model to probe into the semantics of the first combinatorial forms, reveal nothing to us about the performances they satisfy. What are concretely the reasons why the child addresses to his interlocutor a gesture-word combination with agent-object or action-locative roles? Although all communicative signs are motivated by the will to interact, it is through effective and circumstantial acts of communication that the baby manages to refine the possibilities of his exchanges with the adult in such a way that this first will to exchange is broken down into more and more specific motives. Ergo, it must not be forgotten that behind the duples of juxtapositional structures (which allow us to understand how verbal and nonverbal signs are coupled), the form of expression and each content involved respond to specific intentions. In this scenario, the same sign combination could have been formulated from different behavioral aims. If the pointing gesture can be protodeclarative or protoimperative, the combination of 'I' + pointing on some object (or the utterance of the object's name) can pursue an informative purpose, highlighting that/the object is

mine/, or an appellative one (/give me the object/). On certain occasions, the goal of a certain utterance can only be established if the interpreter is aware of the melodic contour, the musicality with which it has been delivered (the interrogative purpose will be evident in the prosody of the verbalization); on other occasions, it will be the facial expression or the bodily attitude that can clarify the meaning of an oral utterance. The *in situ* consideration of the semantic duples or multiples will invariably be that which will provide access to the function behind the form.

In F. G. Rodríguez and Español (2019), we explored the field of different communicative functions. In what follows, it will be first discussed the criterion we have adopted for our observation code. Then a variety of examples of such functions (with single or combined signs) will be given. Finally, some borderline instances of the notion of communication will be examined. These instances contain features of primitive musicality intermingled with sign forms addressed to an interlocutor (discovering the continuity between primary dyadic exchanges and communicative ones). This continuity, readers will know from the previous chapters, is central to the idea of development in the narrower sense: it allows us to appreciate how the resources that structure the first relationship between a bay and an adult endure and intertwine with the subsequent communicative acts to appear, in other words with a more sophisticated type of socio-cognitive exchanges. In the exchanges with signs executed for decoding, the inaugural patterns of intersubjectivity underlie in a lasting way.

8.2 The Functional Classification of Pregrammatical Communicative Expressions

This section provides a concise review of studies on communicative functions in young children (Sect. 8.2.1), presenting their categories and discussing their problems. Then it is reported a single-case investigation conducted (Sect. 8.2.2) and is displayed the observation code developed for communicative functions prior to the emergence of morphosyntactic structures (Sect. 8.2.3).[1]

8.2.1 From Linguistic Functions to Communicative Functions

As has been pointed out, the study of communicative functions has precedents in the exploration of language functions. Of these, the model that Roman Jakobson summarized in his text on poetics (1960), composed of the functions (mentioned above)

[1] As in Chap. 7, some of the observations have been previously published in different contexts (Rodríguez, 2018, 2021) and are now offered in connection with others to delineate an overall assessment of multimodality in the beginnings of communicative semiosis.

informative, conative (appellative in Bühler), emotive (expressive in Bühler), phatic, poetic, and metalinguistic, has remained as a paradigm, a starting point and at the same time the target of assorted criticisms. One may immediately wonder why a research–such as ours–that wants to address the multimodal structure of adult-baby interactions would take into account a repertoire of functions originally conceived for the competent speaker *verbal* expression. The answer is that such functions are not of language, but have simply been recognized and studied in language before other modalities. Work on baby's gestures, vocalizations, and pregrammatical combinations has shown how deeply the purpose of correct sentences actually formulated is already present earlier in simpler and no less communicative sign varieties. A second objection to the adoption of the Jakobsonian model is that it does not allow us to inventory every utterance in an exclusive way. In certain cases, it becomes difficult to distinguish to which category a particular expression is more appropriate. In this regard, it can be pointed out, as the famous linguist had remarked, that there is usually a predominant function (Jakobson, 1960), even taking into account that in every expression there are harmonics or resonances of other functionalities. On the other hand, when this predominance is not sufficiently clear, the expression can always be included in more than one category. In any case, the investigation of communicative functions cannot be inhibited by difficulties of a methodological nature that can be remedied through appropriate clarifications. If, on the other hand, it is assumed that at the origin of communication is the will to achieve certain purposes *by means* of the other and to do things *together with* the other (again: protoimperative and protodeclarative gestures), research on communicative functions is linked to adult-baby interactions, and should have its first focus there.

The cataloging of linguistic functions has proliferated since the old model of Jakobson (1960), and the same has happened in the framework of prelinguistic functions (or pregrammatical, as would perhaps be preferable, since before the child forges sentences according to regulated structures he can speak, even if only half-language, and is therefore within language). Some works on pregrammatical communicative functions concentrate on certain signs: gestures, vocalizations, first words, first combinations of gesture and word. Others deal with some functions without aspiring to cover the whole arc of possibilities. The cataloguing of all the communicative variants of the child awaits the creation of a complete observation code capable of satisfying all specialists. Some attempts in this sense (Sarriá Sánchez, 1991, Rodríguez, 2021, Rodríguez & Español, 2019) are neither fully comprehensive nor multimodal and require completions.[2] In accordance with her multiaxial proposal, Encarnación Sarriá Sánchez explored communicative behaviors including, along with the more obvious and incontrovertible forms, the touching of objects and people, the action of pushing objects away, and eye contact. She divided the whole spectrum of the baby's communicative behavior into the functions:

[2] In terms of completeness, the system of categories for communicative functions developed in F. G. Rodríguez (2021) and F. G. Rodríguez and Español (2019), which is the basis of Sect. 8.2.3, fails to consider all the important aspects of the act of communication. This system is limited, as has been pointed out in the previous chapter, to the oral and gestural modalities.

object demand, help demand, interactive play with and without object, rejection, protest, and declarative expressions (Sarriá & Rivière, 1991). The first two make a distinction within the protoimperative, the third makes room for the ludic dimension, the fourth and fifth correspond to primary types of attitudinal expressions, and the last amplifies the gestural notion of protodeclaratives. The discrimination between different contents of asking (object, help) is replicated in different research (e.g., request for object, action or cooperation, in Muñetón Ayala and Rodrigo López (2011); request for information as a proto-interrogative gesture, in Southgate et al., 2007 and Rodríguez, 2009). All the ramifications around what the baby demands should be integrated, however, in a generic category: that of communications demanding something from another (conative *sive* appellative function). In our observation code (see Sect. 8.2.3), we have disregarded subdivisions for the basic communicative functions, not because they are not important distinctions, but as an ordering criterion to avoid confusion between levels, something that is sometimes overlooked.

A different case is the division of the declarative role in the pointing gesture, for which two variants have been recognized, the declarative-expressive function and the declarative-informative function (Tomasello et al., 2007; Cochet & Vauclair, 2010). The former is embodied in expressions that aim to share an experience with the interlocutor; the latter in others that provide a certain clue that the baby has noticed is missing in the adult. Rather than merging here the two alternatives into a larger category, as proposed for the various conative subfunctions, it is appropriate to make a pragmatic or *what for* criterion prevail, and therefore to separate the pointing gesture centered on referring from the pointing gesture centered on the transmission of information. From the pointing gesture, private functions have also been distinguished (Delgado et al., 2010; Lucca & Wilbourn, 2018), which have also been studied in the set of ostensive gestures (Rodriguez et al., 2009). These gestures, and indeed gestures in general, could fulfill exploratory, contemplative, self-regulatory, or other roles (see, e.g., Español, 2006), but as self-directed gestures they extend outside our field, which concentrates on communicative signs.

Classifications of gestural signs are very numerous, but they tend to be mostly established outside their functional applications. At the boundary between morphology and use, a distinction is regularly made between interactive gestures, made to moderate exchange with interlocutors, and representational gestures, which possess meaning, reference, or both independently of other gestures (Abner et al., 2015). Interactive gestures are not asemantic, but the meaning of their execution is always linked to representational gestures and context having connection or ordering tasks (they are the corresponding, in the gestural modality, of Aristotelian syncategorematic terms).[3] This relative significance has been reason enough for some specialists to consider these gestures as non-meaningful (Crowder, 1996).

[3] Note that the designation 'representational' here encompasses more than in Chap. 7, where it only included iconic gestures, leaving out deictic and arbitrary gestures.

Murillo and Belinchón (2013), starting from the categories of imperative and declarative used in Bates et al. (1975), studied vocalizations as well as gestures and combinations, including a category for uses 'others' without further distinctions (accounting there for refusals, response to demands, play routines and spontaneous utterances). The result confirmed the relationship, "widely described in the literature" (Murillo & Belinchón, 2013, p. 483), between pointing gesture and declarative function, on the one hand, and on the other hand between reaching gesture and imperative function. The aim was not a structural rethinking of the uses of children's expressions, but to evaluate how the different types of signs fit between the traditional categories of function. This bipartition between the imperative and the declarative allows us to appreciate the extent to which these categories enjoy acceptance and currency.

Sometimes, when the different functions of gesture are approached in an attempt to cover all its possibilities, there is some confusion between a functional and a formal criterion. This defect, Kong et al. (2015) point out, is present even in landmark texts such as Ekman and Friesen (1969) and McNeill (1992) (see a discussion in Rodríguez, 2021).

In an overview of head and hand gestures, Wagner et al. (2014) traces a useful inventory of gesture annotation systems to date, outlining, in two columns, function and form. In the list of the functions you find iconicity, metaphoricity, deicticity, emblematicity, shaping or placing, establishment of spatial relations, emphasis, emotion, etc. This enumeration, however, although it reflects the utilities of the gesture *qua* sign with particular characteristics different from those of other signs, does not discover what the gesture alone or in combination can effectively do by means of deixis or iconic representation. The deictic gesture, as well as the deictic word, can play distinct roles in the context of a communication.

Precisely in Kong et al. (2015) eight functions of gesture were distinguished: providing additional information to that supplied by the word, to enhance speech content, providing some meaning absent in the verbal utterance, guiding and controlling the flow of speech, reinforcing the intonation or prosody of speech, assisting lexical retrieval, assisting sentence re-construction, and non-specific function. Apart from the fact that this coding system is exclusively designed to investigate adult gestures during verbal communication, which leaves out the isolated gesture, the functions highlighted are not all communicative nor do they all correspond to the same primary level. On the one hand, primary and secondary functions seem to be mixed. For example, the functions of providing additional information and providing missing information in relation to oral discourse converge in the basic fact that gesture, in both cases, is a vehicle of information and not a manifestation of subjectivity. On the other hand, under the global category of gesture function, different cognitive competencies are included. For example, assisting lexical retrieval is not a communicative function, but an exercise of subpersonal rank oriented to naming. This is not irrelevant when studying in which general capacities the gesture intervenes during an act of communication, but it goes beyond the communicative field.

A recent study on baby gestures (Guevara et al., 2020) provides an update of the functions categories. These were first separated into 'to others' (communicative)

and ‘to self’. The uses of gesture in ‘to others’ functions comprise declarative, imperative, informative, interrogative, phatic, and evaluative functions. The declarative/informative discrimination follows the coordinates of Liszkowski et al. (2006, 2007) and Tomasello et al. (2007). The imperative and interrogative functions are part of what is more broadly referred to as the conative function, which concerns the sign behavior of appealing to the other *sensu lato*. The phatic function is linked to the action of giving and the evaluative function (“lifting an object as if it were a trophy [ostensive gesture] or clapping [symbolic]”−Guevara et al., 2020, p. 269) could perhaps be interpreted, rather than as a genuine intentional sign, from the baby’s perspective as an intended celebrating *ritual action*. In relation to what we are dealing with, this article, which in fact pursues an empirical objective, has the value of providing the reader with a well-organized state of the art.[4]

For language signs, pragmatically oriented authors proceeding from an inductive approach have recognized a wide variety of actions performed by babies and very young children, but this plurality also mixes levels, as is the case with gestures, and could be reorganized according to Bühler’s first discrimination. Dale (1980) distinguished fourteen functions (name; attribute; comment; tenses; request for present or absent objects; actions or permissions; request for information; nonexistence verbalization; reject; denial; affirmation; attention seeking verbalization; greetings and other ritualized forms; and unintelligible). On the one hand, these functions can be regrouped in the three aforementioned (or in other ways); on the other hand, some of the children’s actual utterances could not be counted in any of the categories (“thus unnecessarily swelling the category of ‘unintelligibles’”−Belinchón Carmona, 1985, p. 47). The difficulties therefore lie in how to achieve a template of functions that covers the largest possible number of communicative uses.

With this objective in mind, and in keeping with Searle’s (1969, 1976) theory of speech acts, Ninio and Wheeler (1984, 1988) proposed a system of categories that was comprehensive of both the utterances and the context of their production, but at the same time focused on the types of exchange, particularly child-adult exchanges. The result was a multifaceted inventory organized essentially around twelve categories of speech acts then divided into twenty-three classes according to illocutionary force (Ninio et al., 1994). Among the distinctions of this second group we find ‘Express sympathy for hearer’s distress’, ‘Express positive emotion’ and ‘Express surprise’, or ‘Ask a wh-question’, ‘Ask a yes/no question’, ‘Ask a limited-alternative yes/no question’, and for all these asking options the corresponding ‘responses’. In view of this, it seems legitimate to ask whether we are dealing with different functions or simply with the function of asking adapted to different interrogative contents and adjusted to different forms of questioning (more open or more closed questions). Even the supraordinal category ‘Questions and responses’ and the subordinate categories ‘Request for repeat utterance’, ‘Request/propose/suggest action’ and ‘Ask for permission’, separated respectively under the categories of ‘Demands

[4] From a similar point of view can also be consulted, focusing on ostensive and indexical gestures, Moreno-Núnez et al. (2020).

for clarification', 'Directives and responses' and 'Commitments and responses', can be integrated according to a fundamental principle of the conversational act: in all of them, the other is called upon to get involved, either to give an answer, or to follow an order or to take into account a warning. Likewise, the responses in each of these last four categories can be declarative, informative, or of other functional types. The cuts of this observation code are governed by the double guideline that concerns the *what* and the *how*, the generic thematic of the exchange and the different modes of interaction, but without clearly distinguishing the intentionality or functionality (in both cases abstract) of the signs themselves from the concrete act of communication. Thus, in the categories of the exchange a distinction is made between 'Discussing the non-present', 'Discussing a present event', 'Discussing the fantasy world' and 'Discussing hearer's thoughts and feelings', without grouping these different topics under the common *discussing* and without separating them on another level from 'Showing attentiveness' and 'Directing hearer's attention'. It may be somewhat necessary to dissociate here two different levels that are often superposed in texts on communicative functions. One of them concerns the functions of the signs, their signifying aptitude, and the other the functions of the act of communication as such, the latter understood as the kind of purposes that induce an individual to address an interlocutor, disregarding the content or the specific *what* (whether *this* or *that* action or thing). From this assessment, making the other attend to a certain object is only a function or a capacity of the sign: it can be achieved with gestural or verbal deictics (the gesture of pointing, names and pronouns, among other options). But what is essential in terms of communication is why we point to an object, whether to ask for it or to provide certain information to our interlocutor. Behind his communicative acts, the child intends to interact in one way or another with the partner, and to carry out this behavior he makes use of the pre-established qualities of signs (to locate physically in space, to represent, etc.). The lack of an adequate segregation between these two different functional aspects (the function of the sign as an abstract competence to do one or more things and the function that the sign undertakes in each circumstance, in a context that brings it down from its *a priori* or abstract limbo and that turns it into a sign at the service of concrete purposes in the framework of real communicative exchanges) can diversify the establishment of basal categories in an inconvenient manner.[5]

In the context of language acquisition, an inventory in Spanish and Catalan (Serra et al., 2000) that was formulated for the first one hundred word-types of the corpus Serra-Solé (1995) distinguished three sets of functions: (i) personal, social, and expressive functions (contemplating greetings, rejection-acceptance, expressions for completion of actions), (ii) regulative ones (requirements of different types −not to be confused with self-regulatory uses), and (iii) informative ones (which include both naming and the provision of various data regarding the referent). In this case, the three types of functions do not fully coincide with the Bühlerian ones,

[5] An applied work of Ninio and Wheeler's category system, extended beyond verbal communications, can be found in Rivero (2010) and Rivero García (2012).

since in (i) there is a mixture of phatic aspects, such as greeting, and elements of exchange with the peer that would be, together with (ii), part of the communicative behaviors ordered by the second person. This interweaving ends up revealing that the triad of Bühlerian functions is at least partially problematic, because, for example, not every expression that requires the other as a co-protagonist (as a receiver from whom a reaction is expected) fulfills the same utility at the basic levels. Interacting for the pleasure of interacting and rejecting a request from the other are both communicative functions but correspond to essentially different interests, which could only be homologated by allowing a criterion of grammatical persons to prevail over that of intentions as such.

According to these reasons, the determining of fundamental categories for communicative functions is open to revision and refinement. On the gesture side, McNeill's coding has dominated over others, but often overlooking the fact that in the interest of this researcher, gestures that are not produced synchronously with verbal discourse, which places the focus on adult discourse, and those that the individual does not generate spontaneously (precoded gestures), are marginalized. On the language side, and apart from what has already been pointed out, the categories have almost always been conceived for speech and have neglected multimodal forms of communication.

8.2.2 From an Empirical Study on Communicative Functions in Early Childhood[6]

In F. G. Rodríguez y Español (2019) we thought it wise to start from a classification that was known in different disciplines more directly or indirectly linked to the study of communication (psychology, linguistics, semiotics, pedagogy, philosophy of language, etc.). At the same time, we were careful not to choose an instrument with categories that would equivocate different functional levels (macro and micro degrees, inclusive and included varieties on the same general level) or with some other type of overlapping. The literature of these disciplines mentioned and the double criterion of exclusion and non-confusion between levels led us to consider the aforementioned classification of Jakobson (1960) plus the performative use of language (the conventional heptad). Because of its degree of disclosure, this set of categories is open to debate and allows for the contrast of statistical results between

[6] As in other chapters, observations indicate age using Piagetian nomenclature: 1; 2 (3), meaning one year, two months, three days. Double quotation marks appear for oral utterances and single quotation marks are used to indicate the mention of a term or phrase. The meaning of gestural, oral or bimodal utterances is entered within pairs of oblique lines /xxx/, with inner parentheses/xxx (yyy) xxx/recovering missing elements in the child's utterance, required for its proper interpretation. The absence of one or more phonemes is indicated by an underscore: x_ xx (e.g., _anks/thanks/). The absence of a complete syllable is indicated by a hyphen: x-xx, -xxx, xxx- (e.g.: -nana/banana/).

different fields of research. As we have highlighted, we understand here that Jakobsonean functions are not merely linguistic functions, but extend to the broader space of communication. Anyone can easily think of examples, remembered or imagined, with which most of these functions can be satisfied pregrammatically, without utterances, using only gestures, vocalizations, isolated words or proto-words, different options of bimodal combinations and behavioral resources that in certain circumstances are loaded with meaning.

Yet, if the aim is to deal with all communicative functions and if it has been decided to start from language, why not start from categories strictly conceived to examine the pragmatic side of verbal events? Why not start from a schema configured for the illocutionary register? Austin (1962) provided a taxonomic scheme of five classes of acts: veridictive (issuing a verdict o evaluation), exercitives (tending to act on a course of action or an event in progress–to endorse or stop it), commissives (implying, for the speaker, some responsibility for a previous speech act), expositives (offering points of view, references, and arguments), and behavitives (reacting to the behavior and states or vicissitudes of others, encompassing actions of thanking, congratulating, applauding, criticizing, protesting and challenging, among others–which counter-intuitively links interpellative motions with uses of the phatic function in the Jakobsonean categorization). As Searle later argumentatively formulated, Austin's discrimination would not be at its core about illocutionary acts but about illocutionary verbs (among other deficiencies: "The most important weakness of the taxonomy is simply this. There is no clear or consistent principle or set of principles on the basis of which the taxonomy is constructed", Searle, 1976, p. 8). While taking this first attempt as inspiration, Searle offered us his version, also of five classes: representative, directive, commissive, expressive, and declarative illocutionary acts.

> The most important conclusion to be drawn from this discussion is this. There are not, as Wittgenstein (on one possible interpretation) and many others have claimed, an infinite or indefinite number of language games or uses of language. Rather, the illusion of limitless uses of language is engendered by an enormous unclarity about what constitutes the criterion for delimiting one language game or use of language from another. If we adopt illocutionary point as the basic notion on which to classify uses of language, then there are a rather limited number of basic things we do with language: we tell people how things are, we try to get them to do things, we commit ourselves to doing things, we express our feelings and attitudes and we bring about changes through our utterances. Often, we do more than one of these at once in the same utterance. (Searle, 1976, pp. 22–23)

Without going into details, the following problem arises. The directive and commissive acts, distinguished by the way in which the receiver is involved (respectively, to comply with a mandate or to be benefited by the fulfillment of some promise by the sender), both *contain*, as total acts and in an inescapable way, the interlocutor as a decisive factor. This point, well understood in the so-called conative function, happens to be a structural feature of coincidence between Searle's two classes, overcoming any difference at the level of *what to do* and *how to do* it. Although this scheme replaces Austin's original descriptive approach with a pattern ordered to the basic functions of the speech act, it falls short of the necessary structural constraints that regulate all communication. Our research, although oriented to the pragmatics

of the acts of signification addressed to an interlocutor, did not remain tied to the *pure* illocutionary, but crossed the illocutionary parameter with the coordinates of the formal structure of communicative processes as found in Jakobson's text. Indeed, it was understood that the places and parts of this structural diagram were a setting of primary relevance, to which the different ways of thinking about the motives of the agent-sender in his intended meaningful act had to be accommodated.

Of course, this tentative category system was subjected to an appropriate survey (although with the limitations of having worked with a single case). Before starting the research, which sought to explore uni- and bimodal communications during the period of first word combinations (approximately 19−27 months, here extended to 29 months and also taking into account gestures, vocalizations, and combinations), some communicative functions observed informally were added, others were split or integrated (in accordance with the criterion of homogenizing all functional types at the same level) and others were discarded which, due to their definition, did not seem likely to be found in a child of this age. The observations were made on an adult-baby dyad. The baby, Bruno, is the youngest son of the author of this chapter. The place of the adult in the interactions was taken by one of the parents or by the babysitter, thus trying to prevent stylistic biases caused by the interaction with only one interlocutor. The video recordings, 15 in sum, were made every 20 days at the child's home. Each videotaping lasted 30 min. The sessions used toys (cars, balls, cube games, insertion games, etc.), storybooks, and everyday objects that varied according to the theme of the game and the space in the house where the interaction took place (child's room, living room, kitchen, etc.). In each session, the adult initiated the exchange by proposing some activity (reading a story, playing with toy cars, etc.) after which the baby was left free to act. The typical session took place with the two interlocutors sitting on the floor, facing each other, at a distance of approximately 1−1½ m. Care was taken to ensure that the two bodies were always within the image frame of the camera.

Preliminary observations led us to reexamine the Jakobsonian classification from a strictly psychological and developmental approach, focusing on the logic of the cognitive processes involved in each particular function. Accordingly, adjustments were made to the initial code. The following is a brief account of the most important decisions made on the set of categories of communicative functions adapted from the Jakobsonian linguistic functions.

The traditional declarative, informative, referential, descriptive, or denominative function was split into a declarative function proper, comprising communications in which the referent is verbally alluded to or highlighted by a deictic gesture (or both at the same time, forming a combination of equivalent or complementary type−see Chap. 7), and an informative function proper, in which, more than an identifying act, a predicative contribution on the referent is recognized. Respectively, the child can make a pointing, use a personal pronoun or name the object ('dog') or can, on the other hand, signify a certain feature of the object (/big/, already verbalized, already by opening his arms wide). The difference is important and clearly corresponds to different cognitive processes, albeit in the field of language it seems to have been largely overlooked (despite the fact that it had been properly

distinguished at the origin of language discussions, in the Platonic texts of the mature period and in Aristotle–*onomázein* vs. *légein*).

Along with this, a second function was separated, now from the expressive function and centered on the communicating subject: the *attitudinal function*. This is embodied in the responses to demands or proposals received from the interlocutors, which reveal the attitude or predisposition of the sender in the face of such requirements. While the expressive function involves personal demonstrations that the subject can even deliver without witnesses (although not in cases of communicative exchange), this could never happen in the attitudinal function, because it implies the reaction to that which comes from the interlocutor. This makes it possible to establish a second comparison. If in the conative function it is a matter of the subject demanding involvement from another, in the attitudinal function the vector of interaction is inverted: the sender receives from his partner an interpellation, some kind of exhortation, or simply registers a message to which he cannot remain indifferent. In the first case, the subject engages the other, in the second case he is engaged. Finally, it is necessary to dispel another possible confusion: that of the attitudinal function with the informative function. Even if the child's intervention can in both cases be reduced to a 'yes' or a 'no' these adverbs may not have a truth value (for lack of correspondence with external reality) but may be the vehicle of the personal positioning of the individual being interpellated. The informative function may contain erroneous data (the adult to the child: 'Is the blue T-shirt in your room?'/conjectural answer: 'No'/correct answer: 'Yes'). The content of the expression with informative function can always be corrected by virtue of the objectivity of the fact referred to, while on the contrary a response of the attitudinal function is by principle incorrigible, since it reflects an entirely subjective positioning ('Shall we go to the park?'/'No').

A function called *echo* was also isolated, which, during the earliest stage of language acquisition, comprises repetitions in which the child picks up the last segment of what was said by the adult. It is usually a word or a fraction of a word modulated with varying degrees of correctness. Sometimes the child repeats this word once aloud and then keeps mumbling it to him/herself. There seems to be no reason for this behavior other than to master new sounds or to understand their referential link or meaning. The child seems to be adding a term to his mental lexicon, although on certain occasions this same replicative action was observed with known words. In these cases, the echo was limited to an isolated and loud reproduction, without the sequential whispered reiterations.

Bearing in mind the stage of pregrammatical communications, the poetic, metalinguistic, and performative functions were in principle suppressed from the Jakobsonian Canon. Regarding the first, nothing reasonably suggests that a child with limited skills in the use of signs can forge structures in which it would be pertinent to speak of a poetic or creative vein. If the poetic function implies the use of tropes and rhetorical figures, at least combination and selection (Jakobson, 1960), and that these various mechanisms are also used with the aim of engendering a particular aesthetic experience among the receivers of the message, it seems difficult to claim that such a skill can be realized with the restricted lexicon of a child who

barely combines two or three words. Novelty exists, of course, in the child's expression, in words and gestures, but there is no right for this to be considered an aestheticizing or embellishing exercise regarding certain underlying content. The spontaneous and novel expressions of children of this age do not have to do with a will to formulate in different ways a basic message, but with the effort to externalize, with the available resources, a thought for which the child lacks the correct name. Let's take an example. It was common that towards the end of the research, whenever Bruno detected a paddle sticking out of his grandfather's sports bag, he would pick it up and place it on his body like a guitar. He would move his hand over the widest part as if strumming the instrument and sing invented songs, mostly of arbitrary sounds, humming or intermingled syllables. It can be seen then how the paddle can be transformed into the guitar (or the spoon into a comb—see Chap. 7) by playfully being given the role of this musical instrument. This *functional transplant* is obtained by endowing the present object with the movement and sound of the absent object and placing it on the corresponding bodily space. But is this substitution enough to assume that there is a metaphor here? Did Bruno really want the other to notice in the replacement a plus of significance such as that which, according to the norm, is to be expected of any metaphor, or has he simply made an assimilation between two things of similar silhouette? It is necessary to take precautions against the lurking confirmation bias (against the precipitation of easily finding early metaphors). In any case, if there were a metaphor in this example, because we define it only through the mechanism of replacement, it would not be a metaphor with poetic eagerness oriented to the creation of meaning effects.[7]

Other examples may restrain our tendency to see metaphors in the expression of young children. At around 29 months Bruno made these three statements (which are not in the video recordings, but are written down as scenes from daily life). Sharing

[7] It is to be understood that to speak of a multimodal metaphor, or of "metaphors we live by" (Lakoff and Johnson 1980/2003), suggesting that there is a metaphorizing cognitive lever that orders our elementary experience of the world, responds to an extra-linguistic and extra-semiotic concept of the term 'metaphor'. From this approach, the metaphorical crosses all other expressive-communicative functions, because it is grounded at the level of the most fundamental experience. That there is a level of conversion between different cognitive domains (or modalities) and that this entails a certain platform for the *a posteriori* creation of linguistic metaphors, is a whole topic that cannot be discussed here. It is only asserted, according to the traditional notion of metaphor, that young children, with little lexicon and who speak by means of word juxtapositions, do not possess the capacity to deliberately create metaphors or to aspire to say things in an original way. A metaphor in the strict sense implies that something with a name can be meant in a different way. Consequently, metaphor requires, as a condition of possibility, the alternative of choosing between at least two expressive varieties. When the child designates some object in an inadequate but understandable way, the amount of appropriateness that this name may have, capable of producing surprise and even approval in the adult receiver, does not imply on the part of the sender that there is any will or *savoir faire* in the substitution carried out. The whole phenomenon can be explained, in the vast majority of examples, by means of an assimilation between the misnamed object and other similar ones, to a varying degree, of which the name is known. Therefore, it is not a case that the child proceeds through a sign selection, but, on the contrary, that he/she is obliged to proceed from his high vocabulary restrictions.

with his father the reading of a picture book, the father would point to some object at every turn and ask what it was. On a certain page, the adult pointed to a ship with sails that, amid large waves, was making a somewhat complicated navigation. Bruno explained: "It's a boat that is swimming". On another occasion, while watching a cartoon on TV, he made this remark: "Captain America throws a wheel" (in reference to the circular shape of Captain America's shield). Finally, on one occasion, Bruno approached his father, who was working on the computer. When his father noticed Bruno's genuine curiosity about what he was doing, he could not resist the temptation to start asking "What is this?" and put various desktop objects in front of him. At one point, when showing Bruno a pendrive and instructing him how it should be inserted into the console, he asked, "What is this?" and the child replied "It's a screw for the computer." In all three cases, if we adopt a broad criterion, the three sentences may lead one to say that Bruno has been able to metaphorize, because within his answer there is a sort of, or certain amount of creativity. It is not unusual for adults to be left wondering about the way children refer to things through infrequent descriptions. But in truth, in the answers provided we see something that can be explained without invoking metaphor and by the concept of overgeneralization (the use of a word beyond its accepted boundaries and occurring in the absence of convenient names). Bruno said of the boat that it was "swimming" not because he wanted to generate some nuance for the idea of moving through the water, but because he had no words to describe the movement of the boat. Now, although at this age the lexicon is insufficient, the physical-functional characteristics of things lend themselves to associations that carry with them the names of known objects. Thus, the action of moving to advance in the water, called swimming in living beings, was applied by Bruno in the case of a vehicle. The vocabulary clarifications will come later, when the child understands that to swim is to move in water only if the mobiles are living beings, and that if what moves is a boat, it navigates (*navis*). The difference between swimming and sailing does not concern the notion of movement nor its physical medium (the liquid), but the mobiles, living being/non-living being. The discrimination between the two seems to exist before the first year (Mandler, 2004, 2008) without a verbal correlate. For its part, the assimilation between wheel and shield does not require further comment: everything is there in sight, in the form of the objects. Finally, the screws do not have encrypted information (like the pendrive), but Bruno associated this new object to an already known one because the pendrives are very small pieces, like the screws, and because these are inserted into other things (they are insertable objects, which he calls 'screws' indistinctly).[8]

[8] Although the older brother, Marco, did not participate in the research (outside of occasional intervention), we have records of some of his cognitive-communicative progress that serve to exemplify what we are discussing. At 2 years and 11 months, Marco used to proclaim "Now is the day" or "Now is the night". On a certain family walk, during a dark night, he suddenly saw a brightly lit shop window, pointed there and said with certainty: "There the day". It could be attributed to this expression a creative purpose, which, however, resents the possibility, more clearly, that in his concepts of day and night it is not yet a question of an hourly cycle, of the sun and the moon, but of diaphanousness and darkness. The day exists where it can be seen, even if it happens in the space of a room with lamps on. Are we before a metaphor, before an event of pure semiotic creativity, or before concepts that the child still has to polish?

In short, it was understood that the poetic function depends on the sender having a certain mastery or ability to select among expressive resources. The over-interpretation of parents and researchers has led to find the metaphor in children's graces or mistakes. If the child makes a mistake, the error can be associated with the correct denomination and, in a confirmatory spirit, put on the sender an inventiveness above his real possibilities and communicative purpose. Accordingly, we assumed that it was not to be expected in this instance to find interventions with a poetic function.

Similar reasons led us to eliminate the metalinguistic and performative functions from our observation code. The decision was flexible so that, if cases were documented, the measure of exclusion would be re-examined. In principle, by concept, it seemed difficult that a child beginning to link words or to form bi- or multimodal combinations could inflect an utterance on itself, taking some sign element as its thematic content. At this developmental stage, gestures and words are directed at objects, situations, and characteristics. The split between an object-language and a metalanguage requires syntax skills with which to encompass one plane in a higher one. The intuition whereby we dismissed the possibility of finding such interventions at this age was that rising to a metalinguistic level requires two enabling factors: a vast vocabulary, so that language can talk about itself, and a certain awareness of language so as to distance oneself from it and turn it into content (in other words, an awareness that allows not only to think *with* words but to think *about* words).

In relation to the performative function, this presupposes, in the oral modality, verbs that are actually verbs, not mere words, as is still the case with pregrammatical children; secondly, verbs that when enunciated perform the very thing they are enunciating; thirdly, that they are conjugated in the first person of the present indicative; finally, that the content of meaning has, in the mouth of the enunciator, genuine performative validity. The present indicative is the child's first verb tense, and at the beginning the child uses the first person or the infinitive (which are very different forms in Bruno's original language, Spanish), but verbs do not become conceptually real verbs, distinguished from adjectives or nouns, until they are accompanied by a subject; before that they are only an effective voice in the context where they have been uttered. On the other hand, performative verbs often imply subordinate clauses ('I warn you that X', etc.); where this is not the case ('I congratulate you', 'I forgive you'), the problem remains of assessing whether the child understands what he or she is doing when using them. In addition, these are verbs that are not usually part of the first vocabulary. Under this perception, we started from the idea that cases of performative function would not be found at this stage.

It will surely be surprising that the recognized first communicative function, the performative, which appears in the beginnings of gesturing in the form of protoimperative and protodeclarative, may lack a corresponding space in our classification. Although the child lacks the aptitudes to use linguistic signs in this role, in the matter of gestures, the deictics (pointing, showing, etc.) would be performative. Now, gestures pointing or showing are used both to share (protodeclarative) and to demand-request (protoimperative) or to inform (Tomasello et al., 2007). In this way, performative deixis fulfills different objectives, and these objectives are in fact the

functionalities behind the signs. The performative so called is thus in the declarative function, in the informative function, and in the conative function. This is tantamount to establishing that performativity is not as such a function, but a dimension of the sign. It would be possible to postulate two dimensions of signification, one performative and the other non-performative. If deixis communicates in a performative way, this does not imply that this performative condition is in itself a function (a communicative act with the same rank as providing information or demanding). If performativity is not the motor of any expression, it is therefore not a type of function, but a register in which certain signs operate. It has to do not with a *what for*, but with the very nature of certain signs to enact what they utter. This is the reason why not even deictic gestures are an argument for including a performative function in our observation code.[9]

8.2.3 Observation Code

Our observation code resulted as follows (nominal and operational definitions, examples, and discussion of borderline or controversial cases are included).[10]

Declarative Function Expressions that involve recognition and naming of objects, facts, or properties. The declarative function intends only to orient the interlocutor according to the referent without adding other types of information. The underlying cognitive process is an identificatory act. The subsequent expressive act only seeks to make this content transparent to the interlocutor. The sender uses the signs to mark *this-here* (or *there*). Operationally, the declarative function can be realized by an allusive deictic or representational (iconic) gesture, by different types of oral utterance (a name, an onomatopoeia of nominative−not predicative−purport, a certain non-lexical vocalization that in relation to the context is sufficiently clear as to its designative purpose) or by a certain combination, unimodal or bimodal. In the case of deictic expressions, whether gestural or verbal, the declarative is limited to

[9] Kendon (2004) separates between referential content gestures and pragmatic gestures. The latter have, on their part, different functions. They are *modal*, when they give a glimpse of the sender's position; they concern the partitioning of stretches in the speech act (*parsing* function); they are *interactive* or *interpersonal* (regulating the turns in a conversation or identifying the addressee of a message); or they are *performative*. But performativity is not defined here in our own terms. According to Kendon, "In [some] cases a gesture may have a *performative* function, as when it is used to indicate the kind of speech act or interactional move a person is engaging in. For example, the palm-up-open hand may be used as a way of indicating that what the speaker is saying is 'offered' to the interlocutor, perhaps as an example or as a proposal" (Kendon, 2004, p. 159). But, as has been argued, the performative in this gesture can *function* as part of a subjective emotional manifestation, as a declarative, informative or conative utterance. The fact that performative gestures were by the same author called, in a previous text, 'illocutionary marker gestures' (Kendon, 1995), gives us confirmation that the idea of performativity is taken in a broad sense, different from the one adopted here.

[10] The quantitative results of the research can be found in F. G. Rodríguez and Español (2019).

a localizing act and the sign does not carry any characteristic of the referent. When a name is used, on the other hand, there is a categorization (verbal or representational gesture), which places at the level of the expression certain properties of the alluded entity (for example, by saying 'dog' the dog is distinguished from the dresser, while by saying 'this' the sign does not reflect differences between the two objects). Both the pointing gesture and the nominative verbalization, when they occur in isolation, could lead one to believe that a declarative function unquestionably takes place there. However, both alternatives, if they are, for example, a response, may have an informative function. If the adult were to ask 'Where do we sit?' and the child were to say 'Chair', the noun in this case (and others like it) constitutes an apocopated version of the truth-valued mental structure/we sit on chairs/.

An expression with an imitative gesture can also be declarative. At the age of 1; 9 (23), Bruno was sitting on an armchair and discovered in a corner of the living room the toys not put away from the previous day; a little car with which he had been amused stood out, and looking at his father, he made with an open hand, palm down, the gesture of moving it from one position to another. To confirm he had understood, the father asked, "The car?" to which Bruno nodded his head in the affirmative. When asked if he wanted it, he verbally expressed, "No". He just wanted the interlocutor to see what he had recognized.

Informative Function Expressions with attributive intent and implicitly or explicitly possessing propositional structure (S is P). Unlike the expressive function (see below), which can also have a skeleton of subject and predicate but with fully subjective attributions ('this is horrible'), here the attributions can be ascertained and are, consequently, objective descriptions ('the pencil is red'). As soon as something is objectively true or false, its assertion can be rectified ('the pencil is brown'). If declarative-type expressions consist of the structure of a certain this-here (or there), this function carries the idea of this-that (or so) or this-such (or which).

As in the previous category, the informative function can be realized by isolated signs (gesture or word) or by combinations. The simplest imaginable case of combination is that of a supplementary combination (see Chap. 7), where the gesture points to the object and a verbalization couples a feature (or vice versa). In 1;9 (11), in a lunch situation, Bruno burned himself on the first spoonful of food and spit it out. In view of this, the mother invited him a new mouthful by blowing on it to cool it, but Bruno, distrustful, looked her in the eyes, pointed to the food and said "Burns" (supplementary attribution: /this burns/). The logically bicomponent structure of the attribution may also require only a single sign. In a dialogue situation, the adult could formulate a question, for example, 'Where is the slipper', and the child could perfectly well respond with a localizing-informative sign (recall Tomasello et al., 2007), or with an adverb of place, 'There' (or another type of word, depending on the question). Even if the child's quantitative contribution is only one sign, it is

cognitively ordered with the sign previously provided by the interlocutor and configures a propositional structure.[11]

Conative Function Expressions that actively involve the co-participant as the one from whom a response or reaction is expected. In this sense, this function implies requests, orders, claims, and questions. More than the rest of the functions, it highlights the need of the fellow human being that every person experiences at the fundamental level. The conative (appellative or directive) can be produced with or without interjections, names, or vocative idioms (such as the Argentinian 'Che'). The demanding or commanding tone is evident in the semantics of verbal or gestural signs as much as in prosodic markings (increased volume in imperative requests, special melodic contour in inquisitive uses, etc.).

At the age of 1; 8 (2), during a beach vacation, Bruno was playing on the beach, surrounded by his family, digging with both hands in the sand to make a well. He remembered that he had a plastic shovel that he used every day. He looked around and since he couldn't find it with the naked eye, he stood up and went to where his father was. He stood in front of him and waited for him to notice his presence. When this happened, he pointed to the place where he was digging the well and said "Come on!" in a tone of complaint. The father understood and went for the shovel. Bruno returned contentedly to his task, thus confirming that it was the object he desired. He had only pointed to what he was doing and had employed a term of many uses, leaving in the limbo of the unexpressed precisely the name of the object; but from the silence to which the shovel had been confined, it became evident by context and the adult was able to reinstate it as another element of the message uttered.[12]

[11] Support in the environment corresponds to a basic premise for the effective expressions of this stage. At the age of 18 months and with few words, Bruno's brother, Marco, saw reappearing at the front door, welcomed by his mother, the person who the week before had started to work on the domestic chores. She greeted him, smiling, but Marco only managed to make a noise (a harsh, dull, intense throat sound) that none of the bystanders (father, mother, the maid) could associate with anything. Then he headed for his room, stopping every three or four steps to make sure he was being followed. He stopped before the closet doors, waited for the others, and opened them wide. Then he repeated that mysterious noise and pointed to the vacuum cleaner. Now it was clearer: seeing the maid, Marco had suddenly remembered that in previous days she had cleaned with that appliance. When he first made the noise, the vagueness of the utterance and the lack of context made it impossible to understand what later, through a sign combination, was a reference to the cleaning tool. With the gesture of pointing, the noise−now associated with a particular object− became onomatopoeia without any problem. The person of the maid had evoked in Marco a certain activity, and the two expressions (first the sound and then the gesture-sound combination) were a predication with the structure /she + that artifact/, which without any forcing can be equated/ to / she (used) that artifact/. /She/, because she was present, did not need to be signified. The redundancy between gesture and vocalization, both concerned with the same object by means of different ways of signifying it, was therefore laterally referred to the human figure that triggered the whole process.

[12] The verbalization of this episode is no more than emphatic, but by marking, through gesture, what it was about, the tacit element could be understood with a pointing to the place where the well was being made. By the way, this is a combination that does not respond to the standard categories

Expressive Function Expressions intended to notify a personal feeling, impression, or appraisal. It must be considered that the expressive−or, among different denominations, the emotive−*communicative* function is less than the expressive function *sensu lato*, because the subject can make use of the gesture or vocalization in an environment without other people, or that includes other people without the expression being addressed to them. Accordingly, a unimodal or bimodal combination, an isolated gesture, an isolated word, an onomatopoeia or an interjection fulfill the purpose of communicating when they are formulated, with varying degrees of awareness, for a fellow human being and seeking to initiate or prolong an exchange. The element of expressiveness, which of course is invariably present–in one form or another−in every human manifestation, is here affected to give deliberate public testimony of subjectivity.

A word or a phrase may be informative in terms of a first aseptic description, but when a certain prosody is added to it the same content may procure nuances of anger, contentment, irony, or many other subjective marks. The interjection and the emphatic or emotionally charged gesture are signs naturally affected to perform this function, but the modalizers of language, the subjectivemes, the *pianissimo* or *fortissimo* of the act of emission, the vocal or gestural melody (since also the gestuality puts in view of the observer the affective tenor of the individual) modulate personally the content of the message. In 1; 10 (14) Bruno exclaimed, almost in a shouting tone, “There’s water!”, when he detected a puddle formed on the kitchen floor, at the base of the refrigerator, revealing his surprise. At another time, without the emphasis of this episode, this could simply be a notifying act.

Attitudinal Function Expressions with which the subject positions himself before an intervention of conative function on the part of the interlocutor. Unlike the previous function, here there is no subjectivity displayed motu proprio, so to speak: out of nowhere, but as a reaction to requests or demands coming from outside. The attitudinal function is never a first movement but rather a response to an intervention of the other. Thus, in the cognitive sphere, it is not only emotion that is at play here, but also, in a prevalent way, processes of the propositional attitude type (from which this function takes its name). The mentalistic verbs (to believe, to think, to think, to think, to wish, to hope, etc.) reveal the predisposition or the attitude that a certain idea awakens in the individual (to believe that p/not to believe that p). The expression of the subject is sponsored by something previously manifested by the interlocutor and to which he/she joins or from which he/she departs.

At the age of 1;8 (25), a few days after the summer vacation mentioned above, Bruno was visiting his grandparents’ house with his parents and his brother. He was playing alone in the back garden. At one moment, his father called him to continue the play inside the house. He heard clearly, made eye contact with the father, but did

(as has been stressed in the previous chapter), because the signs are not redundant, nor do they imply, as in the supplementary combination, an attributive compound. In this case, the gesture highlights what is at issue and the word is a sign of exhortation (“Come on!” = /help me/).

not get upset. When he was called again, in a firmer tone, he stared at the father, drew his right arm back, as if retracting it, gathering momentum, and then vehemently thrust it forward, the index finger prominently outlined from the rest of the fist, pointing at the one who was bothering him, and uttered at the same time, very simply, 'No'. The gesture underlined to whom the refusal was directed in the oral modality. The meaning was crystal clear: /no to what you say/.

Before the child can construct subordinate sentences, his or her still limited semiotic capacity is nevertheless quite competent of pronouncing on the contents of another's expressions (suggestions or proposals to be replied to). At the stage of the first combinations the child can verbalize 'Don't want' (whatever), or reject by means of a gesture, or combine both modalities, but it is enough for him/her to administer the adverbs of negation or affirmation or the equivalent gestures. The 'yes' or 'no' that the child expresses as an attitude is indifferent to truth conditions, proper to the informative function.

Phatic Function Expressions that seek only the care of the communication channel. The content of the message matters in them less than the mutuality practice as such (it is hierarchically below the pragmatics at stake). Needless to point out how the protoconversational legacy (from the time when vocalizations and physical forms of contact were the sole support of the adult-baby relation) seeps in here. At a certain point, these marks of origin endure throughout life. Along with contents of increasing semantic complexity, interactions continue in part to run on tracks in which the message is less relevant than the interpersonal nexus.

The phatic function is tied, at this early time, to conventions of the pragmatic level, to words that open and close conversational exchanges, to certain language formulas ('hello', 'excuse me', 'thank you', 'please', 'bye'), to the equivalent gestures and to rituals that must be performed to fulfill learned routines (for example, clapping when a goal is achieved, e.g., raising a tower of cubes). These are canons of initiation, maintenance, and completion of communicative exchanges. All these and other words and gestures for framing communication were not only part of Bruno's sign repertoire, but he was able to use them with ample solvency, without errors, placing them in the appropriate places. Among Bruno's pragmatic skills at the beginning of the period under study, he not only respected the verbal standards of channel maintenance, but, despite his limited lexicon, he used to employ any words, regardless of their meaning, to keep their place in a dialog. At the age of 1;7 (13) he used negation as a crutch. He answered with 'no' to questions that he understood perfectly well could not be satisfied in that way. Because he was focused on something else, he answered without answering, draining the words of their meaning: the 'no' had become phatic for the sake of saving face and making the other feel that, somehow, he was not unaware of what he was being told. At other times, without even bothering to de-semantize a word, he sustained the turns of the interaction with asemantic vocalizations, interjections ("Hmm", "Uh-hu") repeated for no rhyme or reason. In other cases, without taking his eyes off a toy he was playing with, Bruno would respond to the interpellations by raising his shoulders every turn: the adult would speak to him and he would comply, as if by reflex, by performing

this physical act. The adult would insist, come back with another question or comment, there would be a minimum of silence, and Bruno would perform the gesture again as if by automatism.

Echo Function Expressions in which the child repeats what was said by his interlocutor in the previous moment without semantic or pragmatic link to the topic of the conversation. The fact that the repetitions observed were mostly new words leads us to think that the echo is mainly aimed at incorporating them into the child's lexicon. Bruno used to repeat the new terms in the adult's utterances as if he wanted to modulate what he heard and to reaffirm its meaning content. Whenever the interlocutor's utterance consisted of more than one word, the repetition took the last segment (from where the name for this function has been selected). In certain (few) cases Bruno repeated known sounds, perhaps to retain the application of the term to a different context. Of course, neither lexical acquisition nor the use of words in new contexts can ignore the ever-present ludic component, which could in itself serve as an explanation for these repetitions.

Even if these repetitions were related to the expansion of vocabulary, which could indicate that the child speaks for himself, the communicative nature of the echo function cannot be questioned. The repetitions were made while looking at the adult and remaining silent for a few moments, as if to clear a space for the pertinent corrections to be made. After that fraction of time, the expectation ceased, Bruno adopted a more relaxed attitude and seemed to be sure that he had understood and reproduced well what he had heard. On certain occasions, the repetition was altered prosodically and turned into an interrogation. This factor, if there were any doubts about the communicative character of the echo function, dissipates the controversy, since for the echo a previous and triggering stimulus-emission is necessary, and this comes from the others. As much as that, the other is necessary as a receiver of the repetition. The gaze suspended in the adult seems to endow him/her with supervisory attributions: power of veto and rectification for what is badly reproduced. The cases of eco-interrogation should be counted doubly in quantitative analyses (as echo function and conative function), although strictly speaking the echo seems to be subordinated to the intention of clearing the doubt with a response from the adult.

As in the phatic function of this period (where the child's intervention can even be de-semantized), the echoic performance seems to go against the transmission of information and defy the idea of a communicative exchange. The difference lies in the fact that, while in both cases the topic of the dialogue is neglected, the phatic function does so in order to preserve the channel, whereas here the verbalization detaches itself from the dialogue and refocuses on the signic matter of the exchange. In the observations made, the conversation could be about clothes, and if at some point the adult named a garment that Bruno did not know, he lost the thread of what was being said and remained pending for the new term, which he tried to repeat.

Before our study began, Bruno emulated isolated words or words he had amalgamated. If he would hear "This", he would say "This"; if he would hear "Good night", he would repeat: "_oo_night". The echo, however, was confined to the last word in long sentences. When he was aged 1; 6 (18), sitting with his brother in front

of the television, as the latter wanted to inform him what the character on the screen was called ("This is Ben Ten"), Bruno replied only with "-en Ten". The strategy lasted throughout the whole period, especially on words he heard for the first time or noticed for the first time, and of which he often could not mime more than the last two syllables. This even at times when he could string together series of several signs in verbal modality. For example, in 2; 4 (1), when he heard: "Let's drink milk", he replied with certainty: "No, not milk, - - fier! /pacifier/. Milk Ma_co/milk (for) Marco/. In 2; 4 (7) he questioned the brother": "This -puter yours?" /Is this computer yours?/. But at the same time, he was still imitating the new words defectively. At 2;4 (24), an adult instructed him: "This is a racket", and he repeated: "-cket"; "This is an accordeon", "_cordeon". This implied that progress in the elaboration of sentences did not imply discarding the resource of repeating where he encountered the same difficulty as in the early days.

The learning hypothesis for interpreting the expressions of the echo function and the modulation of the words heard in a question tone could suggest that we are dealing with metalinguistic uses. After all, the child repeats in the form of an interrogation and thus enquires about an element of the language. However, it should be noted that the child is inquiring about the relationship between meanings and signifiers or about the proper diction, not about the conceptual contents of what he is saying. Saying '*White* begins with *w*' or asking 'Is *nice* an adjective?' are metalinguistic uses; the prosodic formulation of a question that repeats in echo what is heard (X) rather corresponds to '¿Is this X?' or 'Is this called X?', or to 'Do I say X correctly?'. In these two cases, the pattern of correctness rests exclusively on mimesis and not by grammatical theory. Everything metalinguistic concerns the structure of the language, not the referential usage. Therefore, what the echo gives to language learning is limited to two non-metalinguistic fields: the gymnastics of sound and the semantics centered on the naming act. It is indeed as if, in the echo function, the child was performing *magna voce* what later the working memory will do silently by means of the phonological loop, but at the same time, as if the echo were intended to prolong the life of a certain sign not for a short period of time, but to last in the vocabulary.

Although at this stage we have reported that, within the oral modality, there are only repetitions of lexical items, the precedent of the echo function is to be found in the repetitions that the baby makes in prelexical protoconversation, with which he learns to modulate, as in a mirror image of the adult, the phonemes of his language. Echoic reproduction evolves from the vocalizations themselves to lexical items. Then there seems to be a splitting of this resource into two parallel paths. From one side, hand in hand with the growing understanding of vocabulary, repetition can become a ratifying act (see below) and cease to play the role of mere reflex or rehearsal in order to link sound and referent. From another side, the echo continues in the form of longer and more complex reproductions. An example of the first alternative is when, at the age of 22 months, the father made the following remark to Bruno, at mealtime, showing a box of food to be prepared: "Here is the food inside". Bruno reproduced phonetically wrong: "food-side", but he understood what he was saying and validated the previous paternal affirmation. At a later stage, after the first

mimetic babbling, the echo enters, as we can see, at the service of other specific communicative purposes. A curiosity of this particular repetition is that Bruno has no longer blindly copied the final part of the other's utterance, but has selected those elements that are semantically relevant: *where* and *what* (adverb of place and object). At first, the child's utterance seems to be limited to the final section (recency effect); later he becomes independent of this condition and extracts the meaningful pieces of the total construction. As an example of the second derivation of the echoic function, the following example can be given. At 2;5 (1), Bruno heard: "It does not seem to me", and he reproduced, with limitations: "Not seem to me". But soon after, in the same session, he heard: "It is of the same color", and he immediately replied the whole sentence with impeccable diction: "It is of the same color".[13]

In all the productions in the service of the echo function, it should be noted that, even if they are mostly oral, there are also cases of gestural realization. For example, in 1; 9 (17), while sharing with his mother the reading of a picture book, Bruno made a contact pointing on each of the figures that she had also touched with a pointing immediately before to accompany the names of things. The mother would say 'horse' and point to the corresponding picture on the pages of the book, then name another figure and point again, and so on. Bruno kept silent, but pointed in turn, after the mother, each illustration. In no case did he repeat the names, only the gesture of contact.

This behavior cannot be justified as gymnastics to learn new names, since Bruno knew those names and had opted not to repeat them. Perhaps it was just a matter of being able to understand by doing what the other had identified with name and contact. The following event was registered prior to the beginning of the investigation [1; 4 (19)]. Once again, Bruno was in front of a picture book, sitting on the floor next to his nanny. At the end of the reading, the nanny looked at Bruno and said "Bye-bye" to each of the drawn animals, making the manual gesture of saying goodbye ("Bye-bye, giraffe", "Bye-bye, panther", etc.). Complying with the protocol of proto-conversational exchanges, Bruno also made the same gesture, while observing that the adult had finished before.

Finally, the observation code included two categories for expressions of an imprecise type: the dubious function and the 'other' function. Finally, the observation code included two categories for expressions of an imprecise type: the dubious function and the 'other' function. Although the desire to communicate was evident in both, the pragmatic motivation behind the signs could not be understood, either because of diction defects (in the oral modality) or for some other reason. As already

[13] It is appropriate to recall here the example from the previous chapter in which Bruno transfers from the oral to the gestural modality an element of the other's intervention. The mother had told him: "It's not a cookie, it is a noodle" and immediately clarified "You don't eat it". Bruno copied the message by taking the negation to the gestural modality and preserving at the speech level, as a question, only one word: "Eat it?". This had been interpreted by us as a case in which the total input, perhaps a very demanding verbal production, had been decompressed and split into two modalities for the relief of the sign processing system. The example can here be inserted between the echo function of single final elements (the last word) and the longer reproductions that keep the modality of the input unaltered.

discussed, while the cases where two functions disputed the main role without it being possible to decide which was more important were counted in the two functions indistinctly, the category 'dubious function' gathered those expressions for which not even one function could be clearly established. These were cases in which the signs, within the circumstance in which they had been produced, did not allow any recognizable purpose to be discerned. This can be illustrated with a combination of deictic gesture + demonstrative pronoun: in certain situations, this combination fulfills a declarative function, in others an informative one, and in others it consists of an undecidable communication. Apart from this dubious function, a space was left for gestures, vocalizations, or combinations not foreseen in our categorization (other-function).[14]

8.3 Within Our 'Other' Category

Up to this point, our system of categories has merely taken a traditional code in the field of linguistics and adapted it to an early stage of the child's communication, contemplating not only words but also preverbal vocalizations, gestures, and bimodal combinations. The qualitative review of the other-function put us on the track that for its anomalous cases, at least for some of them, tentative groupings could be formulated. In principle, we managed to distinguish three new types of functions that cannot be assimilated to the categories of the code from which we started. As can be seen, two of these three types lead us beyond semantics, since on certain occasions the signs obliterated their decoding, whether they were gestures or words, and thus relinquished their value as such. In other occasions it could happen that the use of orality or corporeality was *per se* significationless: oral and physical behavior seemed to return to the parameters of the most elementary interactions, where movement and pure sounds prevail, under the guidelines of early musicality. In this way, thus, it goes beyond the frontiers of communication, at least regarding the notion of communication used in the previous chapter and again called upon at the beginning of this one (meaningful intended actions).

Next, the three new function types identified within the other-function are defined.

[14] Tangentially it can be added, in order to make the relevance of these two categories visible in our results, that the other-function and the doubtful function had a high frequency and ranked third and fifth among the total categories. The sum of their two frequencies brings them to second place, only behind the informative function. This was a singular and somewhat uncomfortable fact, as it hinted that some central consideration could have been omitted in the coding. It was the same deficiency that Belinchón Carmona had pointed out in Dale's categories (see above, Sect. 8.2). However, the fundamental categories of all the other codes consulted had been contemplated and others of our own had even been added. Modulation restrictions, coupled with the ambiguity of certain situations, must be counted as partially causal factors in these results. In any case, this problem began to mend when we took a closer look at the interventions within the other-function (see Sect. 8.3).

8.3.1 Ratifying Function

Expressions that support an assertion made by the interlocutor. It differs from the informative function because, unlike the latter, it does not provide new information of any kind, but rather lends an endorsement to what the other has said (an endorsement which, by the way, had not been requested, so that the expression has arisen without being the response to a query). It is not an expressive function either, since there is no spontaneity or manifestation of a free opinion. The child seconds with his intervention something stated by the interlocutor, and does not do so as if it were an echo, but rather by giving evidence of being on topic and of having grasped what was said by the partner. He does not offer a point of view or stance in relation to a proposal received (attitudinal function), but rather supports a descriptive content. Even if the form of ratification sometimes implies repeating literally what was said before by the other, there will only be a ratifying act when there is at the same time a genuine understanding of what is being discussed. As a last distinctive feature, this function cannot be confused with the phatic function because, centrally, what is relevant in it is not the care of the channel, but what is signified.

At different times, and according to his increasing semiotic and communicative skills, Bruno would accompany what he was hearing from the adult by means of gestures or words. It could be with an adverb, used as a corroborative ('Yes'), or with any term by means of which he made his agreement evident. For example,

> Observation 1: 1; 11 (26). The mother asks him what liquid is inside a glass, and he answers correctly, "Water." "There is a lot of water in the glass, isn't there?" the mother then clarifies, and he approves by repeating "A lot" and assenting with his head. Since he usually distinguishes 'a lot' from 'a little', the repeated word has an appropriate semantic and pragmatic meaning. The degree of ratification could have been provided simply by the gesture, but here it takes place by a bimodal combination.

> Observation 2: 2;5 (1). Sitting on the floor in a play situation next to the babysitter, from a scattering of toys around him Bruno picks up a cylindrical can that was originally the wrapping of a doll. As the doll was somewhere else, he did not expect to find any object inside the can. Finding something out of place there, he makes an astonished face. Noticing this expression, the nanny looks inside the can and says: "There is something inside". Bruno confirms: "Yes". She recognizes the object and adds, without looking at Bruno, "Something Mom bought you". Bruno confirms: "Yes". She asks: "What is it?", and he answers precisely: "Go-gos" (the name of the type of doll). With his first two interventions, Bruno validates what the interlocutor says.

> Observation 3: 2; 5 (5). A few days later, the nanny informs the father, who had just come home from work, clearly intending Bruno to hear her: "Bruno ate all the chicken today". Bruno immediately endorses the statement, showing impeccable grammar: "Yes, I ate all the chicken".

The forms vary over time, and expressions of greater complexity appear, but in the three cases referred to, the function remains the same.

Ratification is not necessarily affirmative. Also with negation a statement is assented to. For example, when Bruno was 2; 3 (13), he once heard his mother complaining, because they were in a hurry, that she was sure she had left a jacket

hanging on the back of a chair and now it was no longer there. She said aloud, with Bruno at her side and the chair in between, "I had put the jacket here [slapping the palm of his hand on the back of the chair]. But it's not here any more! It's gone!". Bruno looked at the back of the chair and then said, "No, not here" also nodding his head in denial. Of course, negation, no less than affirmation, can be found in an informative, phatic, expressive, or attitudinal role, or appear within an echoic expression, but it can *also* be used in ratifying acts.

8.3.2 Completive Function

Expressions that serve to accompany perceived actions or facts. On certain occasions Bruno looked at an object or took it in his hands and performed a specific action, attaching the complement of sounds that can be regularly heard that the object makes, either by itself or when it is used. It may happen, for example, that something falls from the table and when it hits the floor, cushioned by the carpet, it barely makes a sound; in these cases, Bruno could replenish the sound of the impact ("Pam!"). The uses of sound correspond to perceptual habits: to the universal fact that most phenomena possess a certain multimodal complexion. This function uses the oral modality and the gestural modality to couple sound or movement where the sedimented perception indicates that they should be. Sonorization does not constitute here a semantic use of the oral modality, but the operation of restoring the missing sound in a play activity. On the gesture side, it is applied to make visible a movement that, according to usage, is associated with objects and actions of objects. In the playful context of the young child's interactions, these effects of reproducing image or sound, like the special effects in a movie, are additions that reality conjures up, as if it also wanted to show all its faces or modalities in the expression. In play situation, these completions add to the imaginary sphere a certain multimodal depth that aids verisimilitude.

> Observation 4: 1; 8 (13). Bruno is playing with a water pistol at grandma's house. With each shot he vocalizes "Pjjjjjjjj!", a sound that he matches in time with the movement of extending his arm towards several targets, followed by a retraction breaking the elbow. The pistol was empty, so it did not imply squirting at the targets of the shot. For his part, the onomatopoeia added nothing to the sufficiently clear action of the game, but since every shot produces some clatter, which he knows well, he must therefore add to his shots the clatter that did not come from the pistol. After a few shots (at the dog, at the birds, at the flower pots in the garden and at family members who cross him), he aims at the father, then sitting a short distance away. The weight of the pistol has tired his arm and he is unable to keep it steady, in line with his target. He lowers his arm and points with his other hand. Now he fires the shots with his empty hand, letting out a "Pjjjjj!" each time. He has replaced the pointing gesture with an object by a pointing with the hand that merges into a representational gesture. In this replacement the onomatopoeia was not affected. For the hand-pistol, as previously for the toy pistol, the missing detonation was restored by the human voice. Double substitution: first, oral-iconic (the sound of the shots in the form of onomatopoeias); then, from playful pointing by means of an object to an empty hand gesture.

Observation 5: 1; 11(6). Bruno is dragging a toy car along the floor and suddenly makes it take off as an airplane. As he moves it in the air he begins to make a roaring engine sound. It's as if the sound of the game brought the game closer to the fact of a real airplane takeoff. The car is more of an airplane when the sound of the engine gets louder.

The same recipe of adding sound occurred on other occasions, but it also happened that the reproduced phenomena increased their degree of reality through gesticulation. More than once, Bruno played at crashing vehicles into each other (miniature trucks, motorcycles, cars, ships, spaceships). At the moment of the collision, he would release the objects he had crashed, would free his hands and, with open palms, would bring them from the inside out, abruptly, as if describing the expansive force that was to be produced. It could sometimes happen, the most normal thing, that he would make an onomatopoeia of explosion, but also that the movement would sketch, mute, on its own, the expected effects. In all these cases, the body showed by means of gesture a fact of reality. What was evoked through the contribution of visual information could never have been provided through sound. It is thus verified that the completive function is enacted in one, in the other, or simultaneously in both modalities.

Observation 6: 1; 9 (15). Bruno picks up a toy motorcycle and accompanies its rolling on the ground with an oral emission that goes from the beginning to the end of the movement. He exclaims a long "Aaaahhh", forcing his throat (as if to make sensitive that it is a rough engine noise).

Observation 7: 2; 3 (5). The father carries Bruno upright with his left arm in front of the medicine cabinet. The father is looking for something on the shelves and Bruno sees behind a bottle the toothpaste of Schrek, the animation film character, with which he and his brother brush their teeth every day. He points to the knob and simultaneously looks at his father. He then places his index finger and thumb in an L-shape and makes the gesture of rubbing them together as if with a toothbrush. As he performs the movement, he produces a sound on the soft palate, harsh and muffled, "Jjjjjj-jjjjjj-jjjjjjjj", which sounds like a saw. The movement is rhythmic and at each change of direction, to one side and to the other, the linked sound is paused.

Insofar as it is isomorphically linked to phenomena, the completive function implies the reproduction of objective features. Onomatopoeias, because of their iconic quality with respect to the referents, are the archetype of oral modality expressions for the completive function (although they are not necessarily associated with it). In fighting plays with his brother, or in combats between dolls, Bruno simulated sounds of blows or pain ('boom!', 'ahhhh!', 'pow!') and performed theatrically, very committed to the movements and sounds he had to reproduce. When two dolls engaged in hand-to-hand combat, Bruno's output reflected the impetus with interjections of exhaustion, grunts of fury, and languishing moans.

The completive function is governed by the purpose of repeating the structural imprint of the phenomena of experience, which appear to perception, in general, as multimodal in nature. This multimodality would then attempt to replicate itself in the expressions with completive function, as if the multisensory force of reality somehow crept into the baby's enunciative and gesticulatory behavior. This is not surprising, because it also occurs in adults.

Unlike what happens with the activity of attunement, where the adult picks up certain amodal qualities from the child's behavior, here it is a matter of how certain modal characteristics of sound and form reclaim each other to compose the integrity of the phenomenal fact. What we can appreciate in this propensity for completeness is why we make a sound that would not be necessary, or emulate with the body a physical feature of something signified, at the same time, by words. The perceptual contents of each modality are interwoven in the phenomena of the world, in our way and possibility of experiencing them, and this is the antecedent so that later, when we express ourselves, we tend to replace the missing parts. The perceptual solidarities that we observe are those that we are later inclined to regenerate by following those vestiges.

In the example of the motorcycle, the long vowel that accompanies the movement of the toy should not be credited as a sign without prior examination. 'Wow', 'meow' and 'mooo' are consummate onomatopoeias, because their degree of iconicity is such that there can be no doubt about the animals to which they refer. But does the hoarse sound "Aaaahhh" refer to the noise of a real motorcycle? Rather, it seems that the movement has been completed with a random oral emission. The emission only had to coincide with the movement of the motorcycle from the start to the stop point. The sound without the motorcycle would indicate nothing; folded to its motion, it is the hoarse sound of the exhaust pipe. Orality brings of itself an element that becomes a sign only through the relationship with other components of the context.

8.3.3 Philo-musical Function

It consists of interventions where the corporeality or the sounds complete each other according to the model of the previous function of completion, but no longer to reflect the reality of a fact, but to organize a communicative experience according to the principles of musicality. Interventions of this type are regulated by the sonorization of movement or by actions to *imaginarize* sounds in space. Unlike what happens with the previous function, the whole completion has here the musical structure with which dyadic exchanges *temporalize*, *qualify,* and *narrativize* intersubjective contacts (see Chap. 2). As we have discussed, the musical bias cognitively metabolizes shared experiences. It is worth paying special attention to the fact that we are dealing, rigorously speaking, with forms. While in the completive function contents are replenished (the toy falls from the edge of the table and the impact against the floor is translated into the oral modality with a vehement 'pum!'), the musical filiation of this new function is seen in that it does not mimetically reflect anything of reality, even where the qualitative characteristics of the sound (pitch, spectrum) are imposed, but organizes the expression in rhythmic-melodic temporal units harmonized between the modalities. This is why it is possible to distinguish between the two functions by appealing to the dichotomy between content and continent. The completive function adds to a concrete expression a feature that is already part of

the referent as such (it respects in its content a given entitative organization); on the other hand, with the philo-musical function it happens, by its affinity with elements of musicality, that it organizes *forms*, configures *motu proprio* the material of movement or of sounds from which it has made use of. It *contains* in *its* forms, in a musical-creative way, a given experience.

> Observation 8: 1; 8 (3). Bruno turns the pages of a storybook and repeats, over each individual action, an imprecise interjection. The sound is vowel-like at first, a sort of 'æ', then ending in an aspiration 'æj' and finally reduced to the velar sound 'jjj'. The variation in the oral modality kept at all moments the format of the timing of the action. One utterance for each sheet movement. Similar examples in which an iterative action receives a reinforcement of orality were documented at very different times during the observation period, ranging from the beginning to the end, when Bruno could already elaborate multi-word grammatical sentences.

> Observation 9: 2; 2 (29). Bruno was putting the pieces of a concluded game (magnetized letters) in their corresponding box. At the rhythm with which he introduces each letter he makes a vocalization: "Ti" (the first time), later "to-to-to…". Then the mother asks him if he has finished and he gives a negative answer. This "No" interrupts, for an instant, the task that was occupying him. When Bruno takes up the task again, he replaces the first meaningless vocalizations with the adverb of negation, now de-semantized. The word becomes an interjection and each subsequent 'no' marks each action of keeping a letter.

> Observation 10: 2; 5 (1). Bruno picks up from the floor one by one a set of markers to put them in a jar where they are usually kept. At the exact moment he picks up the marker from the floor, he says, in each case, "Tan!", and a sequence is built up: "Tan! Tan! Tan! Tan!". If this sound had been pronounced at the moment of releasing the markers, the expression could have a completive sense, because it would correspond to the impact against the base of the jar; but as each "Tan!" coincides with the action of grasping the markers, which is in itself devoid of any sound, the reason for these vocalizations has to be thought of as part of a structuring of the sequence in sound-effectual units, thus with a temporalization criterion. For the same reason, the "Tan! Tan!" are not in themselves onomatopoeias. Either they are inventions resulting from circumstance, and any other sounds could have been used, or they are the memory of effective sounds once heard, now detached from their sources and applied to soundless events.

In observations 8 and 10, the oral emission only wants to underline, in a certain action, its pulsating structure, the times of the movement in progress. That the same mechanism is observed in observation 9, where the 'no' removed from its role of negation is aligned with the philo-musical function to intermodally emphasize a certain pulsation. With the cautions implied by having worked only with examples, it can be ventured that signs are susceptible to be reabsorbed in asemantic roles such as the philo-musical configuration of subjective experiences. This philo-musical function, with precedents in the most primitive dyadic interaction, does not recede or disappear with the emergence of other uses of orality. The strategies of musicality endure.

In some other cases, silence is the used maneuver:

> Observation 11: 2; 4 (10). Bruno repeats to the interlocutor, who has not understood his pronunciation, some words which, for clarity, he has decided to separate into syllables ("Fire", "Pen-cil"). At the same time, each particular syllable is accentuated by a sagittal movement of the head. With a forward bending of the neck and looking at the adult, he seems to

> indicate to him that each of these voices has the number of parts he is signaling. Verbal separation into syllables does not need this pulsating movement. Something else, the natural bias toward multimodal musicality, induces that segmentation in the oral modality receives an adequate gestural reinforcement, unnecessary but pertinent.[15]

By the use of silence, vocalizations, or gestures, what is clearly seen is the use of multimodal-musical patterns to elaborate daily life experiences. Let's consider the following example:

> Observation 12: 1; 11 (10). Bruno and his nanny are sitting in the park in front of the kindergarten of their brother, Marco, whom they are waiting for. Bruno plays amongst the roots of an old ombú tree. When he gets tired, he sits down next to her. At that, a dog barks three times exactly, woof-woof-woof. She and Bruno watch it. The dog barks again, again for three times. She looks at Bruno and imitates the bark: woof-woof-woof. The dog barks again about three times. A game begins: every time the dog barks, always in the form of three woofs woof-woof-woof, she hits with both palms on her knees, also three times, trying to accompany the bark. Bruno imitates her, a bit delayed, more alert to her than to the bark. He soon realizes that this hitting on the knees coincides with the barking. Then he takes his eyes off her and looks at the dog. When the dog finally barks only twice (woof woof) and Bruno follows the pattern of hitting three times on his knees, he notices the mismatch. He looks quizzically at the nanny. The symphony is broken. Bruno had used a movement module (provided by the adult partner) to match an exogenous sound. He had recognized a pattern and began to comply with it. Then reality (the dog´s right to unexpectedly change its bark), confounded Bruno, who had continued to perform the same metric. Here we see a philo-musical function where the inertia produced by the iteration led to surprise when the hitting and the barking decoupled. The philo-musical accompaniment of the barking had been elaborated together with the adult partner, or rather it had started from him and had been transformed into a common game. That is why, on being interrupted, Bruno immediately looked to the nanny for an explanation.[16]

It is not only in the form of the pulse that we find scaffolding of early musicality. Examples can be cited concerning the spectrum or tonal space (see Chap. 2). A movement can be paired with a melodic oral line that repeats its form in contour, intensity, or duration, thus producing a redundancy of an asemantic type.

> Observation 13: 2; 1 (8). Bruno plays in his room with a toy airplane. He walks around the room carrying it back and forth in his right hand and suddenly decides to let the plane make a swoop from the head of the bed to the floor. The movement was in slow motion, letting us hear, from beginning to end, a long drawn out "Iiiiiiiiuuuuuuuuu" (♫), a sung passage in which the sound descended also, like the airplane, from sharp to low. The sonorization was not the sound of an engine, but rather a melodic contour temporally paired with the up-down trajectory.

In brief, the mechanism underlying these behaviors reveals that the child's mentality is forged in the matrix of multimodality (from which originates the tendency to produce, simultaneously, vocalizations or words and gestures or movements) and musicality (from which originates the tendency to produce temporal and

[15] Bruno's second Spanish word was "Mo-to" (motorbyke), here replaced to maintain the bisyllabicity.

[16] This episode was observed by the father, who had also gone to pick up Bruno's brother from kindergarten and was able to see it all first hand.

non-temporal−quality and narrativity−resonances). The cooperation between these two particular dimensions of the baby's life is a precondition that marks−indelibly− the ways of interacting, something that is projected onto all subsequent interactions of the competent adult speaker. The emergence of the communicative exchange itself is an event that, therefore, could not affect, disturb, or surpass it.

This consortium of multimodality and musicality is seen in certain hybrid cases that meet the requirements to be categorized as both a completive function and a philo-musical function.

> Observation 14: 2; 4 (16). Bruno comes to report that the brother is in the room crying: "Mommy, Marco cry like this: ho-ho…" in a sad voice. The imitation of crying is bodily and sonorous. Each small interjection of crying ('ho') is performed with a movement of raising and loosening the shoulders, resembling a cry of grief. The mother, who at the time of the notification was attentively focused on something else, asks him, "What?", and Bruno, in his response, theatricalizes the information once again. He says "Marco cry", but altering the concomitant illustrative action. He now brings his hand, clenched into a fist, up to his eyes and matches each little 'Ho-ho' groan with a little flick of the wrist (fake rubbing). We are dealing with a challenging case of our categorization, because vocalization and movement are completive aspects of each other, but the pacing of the whine and the flick of the wrist speaks in favor of an expression of philo-musical function. At the same time, we even have elements of an informative function (what happens is true or false). The correct chaining of the words, according to the conventional order of Spanish (vocative, grammatical subject, verb and adverb) represents the particular stage of language that Bruno has reached, but the synchronized effectuation of gesture and onomatopoeia is an archaic resource.

A similar episode, in terms of the functionally half-breed character, has been previously mentioned; it occurs when Bruno, at his father's side and facing the medicine cabinet in the bathroom, mimes brushing his teeth by putting his hand in L, moving it to one side and the other and making a saw-like sound. Now that we have introduced the so-called philo-musical function, the case previously cataloged as completive, insofar as the sound was added to the movement of the brush-hand, reveals new edges. When we focus on the fact that the frictional rubbing separates the sounds respecting each traction−from left to right and vice versa−with appropriate pauses in between, we realize we face a multimodality case where there is musical concordance in the affected elements. The profound closeness between multimodality and musicality puts us in the predicament of being unable to decide which is the main function here.

The philo-musical function traces with movements the profile of sounds and provides sonority to the actions and gestures of the body. If in expressions of the completive function the physical processes lacking sonority receive certain ad hoc additions taken from reality, in expressions of the philo-musical function the emitter does not replace something missing (of which there may exist, by experience, a foreknowledge), but instead proceeds according to amodal perceptual patterns. The philo-musical structure rests on the experience of social play and on the way in which interactions are organized within it according to a temporal criterion. This unnecessary procedure of musicalization has, apparently, no other explanation than the latent and compulsive propensity to endow any behavior of expressive type, not exclusionary communicative, with a certain multimodal density of rhythm or melody. The interactive musicality operates then, through its characteristics, as an

organizing device of the subject's experiences even when gestures and speech have arrived and colonized the dyadic exchange.

The relevance of having defined three new functions, exhumed from the set of expressions of the other-function, can be measured in relation to the total amount of interventions of the child with other-function during the period (N = 613). The summed frequencies of the ratifying function, the completive function, and the philo-musical function amounted to N = 114, which is almost one-fifth of the utterances that could not be classified. The rest of the utterances could not be placed under other organizing principles. The opportunity remains open for the future to identify conceptual clusters for expressions that, within the framework of children's communicative exchanges, have not yet been catalogued.

8.4 Final Remarks

A category system for communicative functions in children with pregrammatical speech has been presented. It was intended to cover all variants of expression: words, preverbal vocalizations, interjections, onomatopoeias, gestures, and combinations between both modalities. We started from traditional categories in the field of linguistics, probed which ones could be observed in the skills of the child during the two-word strings period, and added some of our own that classified otherwise incomprehensible expressions.

Among the novelties provided, an important division must be made. The echo function, at least partially, seems to be associated with the exercise of incorporating terms and to play a transitory role. It is not expected that the adult repeats what he/she has heard from another adult, even when confronted with an unfamiliar word. Unlike all the other functions, which remain present in the adult's expressions once they have appeared, the echo function disappears.

Certain communicative functions (expressive function, conative function, and philo-musical function) seem to have very remote antecedents in prelinguistic manifestations: they arise before speech and gestures and subsist throughout life together with the varieties that appear during development. The child's spontaneous expression gradually adapts to standardized signs, as do invocations or calls through nonspecific resources (shouts and cries), whereas the capture of the musical structure in the adult's performances and attunements are capitalized within the communicative space. In summary, if certain functions refer us to a stage that precedes the acts of communication, others, born from the possibilities of communicating, become consolidated and continue indeclinably throughout life, and others (one for the moment: the echo function) satisfies a concrete need and then withdraws.

Some specifications were made on previously recognized functions. The separation between declarative and informative functions reflects the two essential capacities of all languages, naming and saying (or predicate). Interestingly, this separation is not present in the categories of the most current canon (perhaps because language has remained essentially associated with syntax and thus with predication). The

conative function during this period under investigation makes it possible to link the use of protoimperative and protointerrogative gestures with the speech and gestures of the grammatical moment. The expressive function of this stage makes it possible to appreciate the replacement of mere vocalizations and bodily expressions of a non-specific nature by verbalizations and concrete gestures. Regarding the attitudinal function, it shows this same subjectivity, but in the light of the proposals addressed to the subject (towards which the person adopts a stance). If in the expressive function the content of the expressions has been gestated in the subject, in the attitudinal function it comes to the subject from the partner, so this is ultimately a reactive function. As for the phatic function, it corresponds to the routines within which the child learns to use signs as part of a ritual. The child says 'thank you' as an effect of a practice that he has observed many times and has incorporated into his culturally correct behaviors. In that same circumstance, it is expected that 'thank you' will always be said. To do so provides a favorable disposition on the part of the interlocutor. This mechanism lies somehow at the basis of all communicative function and promotes the learning of any kind of sign, words, amalgams, gestures, or small phrases that the child begins to use without grasping at the beginning any other sense that the situational pertinence (ground of the later more precise meaning of the language signs).

In relation to the functions identified within the other-function, it should be clear that, with the exception of the first, the ratifying function, the completive and philo-musical functions are *liminal uses* of communication, because although their expressions are directed toward the other (the gaze is present to make this evident), they are not always signic in a strict sense and refer rather to the phenomenological experience. If the completive and philo-musical functions lack the signic (semantic) element as conceived for the more restricted definition of communication (communication *about* something), in any case all the abovementioned examples took place in a face-to-face situation, and therefore correspond to expressive variants in the service of communication. This is the reason why it is interesting to highlight them in this type of research. In the completive function, environmental noises are imitated or the individual performs gestures that copy the reality of facts. In the philo-musical function, sounds or body movements are aimed at ordering in a musical way a subjective experience (which occurs in contexts of dyadic exchange). Both functions possess and do not possess, at the same time, properties of the idea of communication as an intentional, intended, and signic act (see Chap. 7). Intentionality as *aboutness* does not seem to be behind sound-kinesthetic completions and modelizations.

Insofar as the expressions of these functions are pure forms, without a content, what is at stake in them are *forms of vitality*. Goal intentionality is present, because the sender expresses himself in them and invariably seeks the attention and response of the interlocutor (he is attentive that the other is attentive), but they lack content intentionality in the classical sense. Nevertheless, in the spirit of an integral ecological code of communicative actions *sensu lato* (such as the one defended in the previous chapter), we see that in interventions of this type, skills of early musicality and multimodality are linked with playful signs and behaviors of all kinds. These skills,

which preexist the communicative ones, are later intertwined with them. Consequently, it seems prudent to establish the distinction between communicative functions, which include expressions that fully comply with the triple requirement of communicative signs, and precommunicative functions, which include expressive interactions directed towards the other but without a thought that was codified. By these terms it is intended to emphasize the link between what endures, from previous interactions, and what is new, which gathers and metabolizes previous competencies. Somehow, certain functionalities exist in baby's interactions and are then prolonged in communicative interactions (so that the former can reasonably be called *precommunicative* functions).

The proposed continuum between communicative and precommunicative functions is consistent with Chap. 7 with respect to focusing developmental psychology research on process rather than outcome. To understand cognitive (in this case, communicative) development, nothing can replace an approach that makes it possible to record precisely the development (rather than deduce it *a posteriori* from chronological cuts or according to a *partes extra partes* sign type dissection). Direct observation reveals the way in which the most incipient forms of interaction, characterized by musicality and multimodality, subsist later in advanced degrees within the exchanges of the child who gestures, who speaks half-language and who finally orders words following the conventional patterns of his language. This complex transition has shown how communicative functions are intertwined with earlier dyadic interactions. The linguistic functions, from which an observational code began to be elaborated, were traced back to their original sources in prelinguistic (not only pregrammatical) communication, that is to say: they were traced back to the enabling structure of any sociality, the cooperative bond between a baby and an adult.

References

Abner, N., Cooperrider, K., & Goldin-Meadow, S. (2015). Gesture for linguists: A handy primer. *Language and Linguistics Compass, 9*(11), 437–451. https://doi.org/10.1111/lnc3.12168

Austin, J. L. (1962). How to do things with words. In J. O. Urmson & M. Sbisà (Eds.), *The William James Lectures delivered at Harvard University in 1955*. Clarendon Press.

Ayala, M. A. M., & López, M. J. R. (2011). Functions of the pointing gesture in mothers and their 12 to 36-month-old children during everyday activities. *The Spanish Journal of Psychology, 14*(2), 619–629. https://doi.org/10.5209/rev_SJOP.2011.v14.n2.10

Bates, E., Camaioni, I., & Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill-Palmer Quarterly, 21*, 205–226.

Belinchón Carmona, M. (1985). Adquisición y evaluación de las funciones pragmáticas del lenguaje: un estudio evolutivo [Acquisition and assesement of the pragmatic functions of language: An evolutive study]. *Estudios de Psicología, 19–20*, 35–49.

Bühler, K. (1934/1979). *Sprachtheorie* [Teoría del lenguaje (J. Marías, trad.)]. Alianza.

Cochet, H., & Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months: Different functions, hand shapes and laterality patterns. *Infant Behavior and Development, 33*, 431–441. https://doi.org/10.1016/j.infbeh.2010.04.009

Crowder, E. M. (1996). Gestures at work in sense-making science talk. *The Journal of the Learning Sciences, 5*(3), 173–208.

Dale, P. (1980). Is early pragmatic development measurable? *Journal of Child Language, 7*, 1–12.

Ekman, P., & Friesen, W. V. (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. *Semiotica, 1*, 49–98.

Español, S. (2006). Possible precursors of private speech: Deictic, symbolic and aesthetic self-directed gestures. In: I. Montero (Ed.), *Current research trends in private speech. Proceedings of the first international symposium on self-regulatory functions of language* (pp. 121−124). Ediciones de la Universidad Autónoma de Madrid.

Guevara, I., Moreno-Llanos, I., & Rodríguez, C. (2020). The emergence of gestures in the first year of life in the Infant School classroom. *European Journal of Psychology of Education, 35*, 265–287. https://doi.org/10.1007/s10212-019-00444-6

Jakobson, R. (1960). Linguistics and poetics. In T. A. Sebeok (Ed.), *Style in language* (pp. 350–377). MIT Press.

Kendon, A. (1995). Gestures as illocutionary and discourse structure markers in southern Italian conversation. *Journal of Pragmatics, 23*, 247–279.

Kendon, A. (2004). *Gesture: Visible action as utterance*. Cambridge University Press.

Kong, A. P. H., Law, S. P., Kwan, C. C. Y., Lai, C., & Lam, V. (2015). A coding system with independent annotations of gesture forms and functions during verbal communication: Development of a database of speech and gesture (DoSaGE). *Journal of Nonverbal Behavior, 39*(1), 93–111. https://doi.org/10.1007/s10919-014-0200-6

Lakoff, G., & Johnson, M. (1980/2003). *Metaphors we live by*. University of Chicago Press.

Liszkowski, U., Carpenter, M., Striano, T., & Tomasello, M. (2006). 12-and-18-months-olds point to provide information for others. *Journal Child Language, 7*(2), 173–187.

Liszkowski, U., Carpenter, M., & Tomasello, M. (2007). Pointing out new news, old news and absent referents at 12 months of age. *Developmental Science, 10*(2), F1–F7. https://doi.org/10.1111/j.1467-7687.2006.00552.x

Lucca, K., & Wilbourn, M. P. (2018). Communicating to learn: Infants' pointing gestures result in optimal learning. *Child Development, 89*(3), 941–960. https://doi.org/10.1111/cdev.12707

Mandler, J. M. (2004). *The foundations of mind: Origins of conceptual thought*. Oxford University Press.

Mandler, J. M. (2008). On the birth and growth of concepts. *Philosophical Psychology, 21*, 207–230. https://doi.org/10.1080/09515080801980179

McNeill, D. (1992). *Hand and mind. What gestures reveal about thought*. The University of Chicago Press.

Moreno-Núñez, A., Rodríguez, C., & Miranda-Zapata, E. (2020). Getting away from the point: The emergence of ostensive gestures and their functions. *Journal of Child Language, 47*(3), 556–578. https://doi.org/10.1017/S0305000919000606

Murillo, E., & Belinchón, M. (2013). Patrones comunicativos multimodales en la transición a las primeras palabras: Cambios en la coordinación de gestos y vocalizaciones [Multimodal communicative patterns in the transition to first words: Changes in the coordination of gestures and vocalizations]. *Infancia y aprendizaje, 36*(4), 473–487.

Ninio, A., & Wheeler, P. (1984). Functions of speech in mother-infant interaction. *Final Science Report to the United States-Israel Binational Science Foundation (BSF).*

Ninio, A., & Wheeler, P. (1988). *A revised manual for classifying verbal communicative acts in mother-infant interaction*. Coding manual distributed by the Department of Psychology, Hebrew University, Jerusalem, Israel.

Ninio, A., Snow, C. E., Pan, B. A., & Rollins, P. R. (1994). Classifying communicative acts in children's interactions. *Journal of Communication Disorders, 27*(2), 157–187.

Rivero, M. (2010). Maternal expression of communicative intentions and pragmatic fine tuning in early infancy. *Infant Behavior and Development, 33*(4), 373–386. https://doi.org/10.1016/j.infbeh.2010.04.002

Rivero García, M. M. (2012). Form-function rules in the first words stage: A longitudinal study of two children. *Anuario de Psicología, 42*(1), 65–84.
Rodríguez, C. (2009). The "circumstances" of gestures: Proto-interrogatives and private gestures. *New Ideas in Psychology, 27*(2), 288–303. https://doi.org/10.1016/j.newideapsych.2008.04.002
Rodríguez, F. G. (2018). La pragmática de la comunicación antes del lenguaje [The pragmatics of communication prior to language]. *Revista IRICE, 35*, 95–128.
Rodríguez, F. G. (2021). *Semántica y pragmática en la comunicación del niño pregramatical. Gesto y palabra antes de la articulación morfosintáctica* [Semantics and pragmatics in the communication of the pregrammatical child. Gesture and speech before morphosyntactic articulation]. Teseo.
Rodríguez, F. G., & Español, S. (2019). Communicative functions during the pre-grammatical stage: A case study/Las funciones comunicativas durante la etapa pregramatical: un estudio de caso. *Infancia y Aprendizaje, 42*(2), 413–458. https://doi.org/10.1080/02103702.2019.1585654
Sarriá, E., & Rivière, Á. (1991). Desarrollo cognitivo y comunicación intencional preverbal: un estudio longitudinal multivariado. *Estudios de psicología, 12*(46), 35–52.
Sarriá Sánchez, E. (1991). Observación de la comunicación intencional preverbal: un sistema de codificación basado en el concepto de categoría natural. *Psicothema, 3*(2), 359–380.
Searle, J. R. (1969). *Speech acts*. Cambridge University Press.
Searle, J. R. (1976). The classification of illocutionary acts. *Language in Society, 5*, 1–25.
Serra, M., & Solé, R. (1995). *Childes database. Romance language corpora*. http://childes.psy.cmv.edu.
Serra, M., Serrat, E., Solé, R., Bel, A., & Aparici, M. (2000). *La adquisición del Lenguaje* [Language acquisition]. Ariel.
Southgate, V., Van Maanen, C., & Csibra, G. (2007). Infant pointing: Communication to cooperate or communication to learn? *Child Development, 78*(3), 735–740. https://doi.org/10.1111/j.1467-8624.2007.01028.x
Tomasello, M., Carpenter, M., & Liszkowski, U. (2007). A new look at infant pointing. *Child Development, 78*(3), 705–722. https://doi.org/10.1111/j.1467-8624.2007.01025.x
Wagner, P., Malisz, Z., & Kopp, S. (2014). Gesture and speech in interaction: An overview. *Speech Communication, 57*, 209–232. https://doi.org/10.1016/j.specom.2013.09.008
Wittgenstein, L. (1953). *Philosophical investigations*. Basil Blackwell Ltd.

Fernando G. Rodríguez is PhD. in Educational Sciences from Universidad de Buenos Aires. He has carried out research and training stays in Germany (DAAD) in Linguistics and Semiotics. He is currently Professor of Semiology and Human Communication. He is member of the Research Group of Socio-Cognitive Processes in Infancy and Early Childhood at the Psychology of Knowledge and Learning Department (FLACSO, Argentina). As a researcher, his work is concerned with the development of communicative skills in young children, the structure of semiotic processes in general and the relationships between mind, sign and corporeality. Author of *Semántica y pragmática en la comunicación del niño pregramatical. Gesto y palabra antes de la articulación morfosintáctica* [Semantics and pragmatics in the communication of the pregrammatical child. Gesture and speech before morphosyntactic articulation] (2021) and co-editor of *Los signos del cuerpo. Enfoques multimodales de la mente y el lenguaje* [The signs of the body. Multimodal approaches to mind and language] (2018). His publications can be found at https://www.aacademica.org/fernando.gabriel.rodriguez

Chapter 9
Forms of Vitality Play

Silvia Español, Mauricio Martínez, Mariana Bordoni, Rosario Camarasa, Alicia Nudler, and Santiago García Cernaz

Abstract Ever since the pioneering works in psychology, play has been considered a hallmark of childhood. In this final chapter, we present a non-figurative type of play that has recently caught our attention. We have named it "forms of vitality play", in acknowledgment of Daniel Stern's forms of vitality concept. Stern posits that forms of vitality are features shared by early social play and time-based arts. We suggest that forms of vitality play can be considered an ontogenetic bridge between them. In early social play parents model forms of vitality through the repetition and variation of their sound and movement, while the baby is an involved recipient who participates with general social behaviors. During the third year in the baby´s life, early social play turns into forms of vitality play: adult and infant play together with

S. Español (✉)
Latin American Institute of Social Research (IICSAL), Latin American Faculty of Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET), Buenos Aires, Argentina
e-mail: sespanol@flacso.org.ar

M. Martínez
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

M. Bordoni
Department of Psychology of Knowledge and Learning, Latin American Faculty of Social Sciences (FLACSO), Buenos Aires, Argentina

R. Camarasa
Faculty of Psychology and Human Relations, Universidad Abierta Interamericana (UAI), Buenos Aires, Argentina

A. Nudler
Laboratory of Research in Education of Teachers, Universidad Nacional de Río Negro, Bariloche, Argentina

S. G. Cernaz
National Scientific and Technical Research Council/Laboratory for the Study of Musical Experience, Faculty of Fine Arts, Universidad Nacional de La Plata (CONICET/LEEM-FBA-UNLP), La Plata, Argentina

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7_9

forms of vitality, both create sound and movement motifs and embellish them by varied repetition. Variations of motifs are carried out in the spatial, energetic, and temporal dimensions of sound and movement. Our aesthetic perspective – adopted throughout this book – has led us to gain new insights into children´s ability to improvise playfully in interpersonal contexts. Through that perspective, we have also been able to observe that children often combine forms of vitality play with symbolic play.

Keywords Form of vitality play · Symbolic play · Early social play · Repetition variation · Sensory motor play · Aesthetics

9.1 Introduction

Play emerges in the animal kingdom, especially among mammals. During their infancy, the offspring of mammals are protected and cared for by parents and vital energy flows through them, so they play and romp peacefully. Play flourishes wherever there is a period during which others meet feeding and hygiene needs and provide attachment experiences. Among primates, the period of immaturity extended like an *evolutionary wave* until reaching the human species, which shows the maximum extension span (Gómez, 2004/2007). As Jerome Bruner (1972) noticed decades ago, the prolonged immaturity of the human offspring is the most relevant biological fact for psychological development. Due to their immaturity, babies require an adult who can get involved and care for them over a long period of time; the extended infant-mother interaction includes now a much larger element of play between them, often initiated by the mother. Concomitantly extended immaturity transformed the bond between adult and infant and paved the way for the development of complex *intersubjective experiences* (Rochat, 2000/2004; Dissanayake, 2001). This longer period of immaturity also allowed plenty of time for the sensorimotor exploration of one's own body and the environment, favoring the evolution of an embodied sense of self (Neisser, 1993) and the acquisition of the erect position and walking (Bainbridge Cohen, 2012; Feldenkrais, 2011). The conjunction of these processes – the extension of playtime in an enriched context of intersubjectivity and the increase of opportunities for self-exploration – allowed play to flourish and adopt differentiated forms that have been named and studied over years of research.

Since the pioneering works of Karl Groos, Jean Piaget, and Lev Vygotsky, play has been considered a hallmark of childhood, enabling children to learn the skills necessary for successful functioning in adulthood (Pellegrini & Smith, 1998). In this chapter, we will not discuss all the different types of play children engage in, but will instead focus on a type of play that has recently caught our attention. We have named it *forms of vitality play*, in acknowledgment of Daniel Stern's forms of vitality concept (Español et al., 2014, 2015). Forms of vitality play is directly related to *sensorimotor play* and *early social play*, and it is also linked to *pretend play*. We

will present a brief description of these types of play, which have been widely studied in developmental psychology. After that, we will delve deeper into the concept of forms of vitality. Finally, we will present forms of vitality play, a type of play that we were able to detect thanks to the aesthetic perspective of development characteristic of our work, reflected throughout this book.

9.1.1 Play in Human Infancy

Whether a certain activity is considered play or not has been and continues to be subject of much debate (Pellegrini, 2009, 2013; Burghardt, 2011). The primacy of means over ends and the non-functional orientation are two constitutive features of play that some consider sufficient to characterize it (Fagen, 2011). Along the same lines, Garvey (1977) described human child play as a pleasurable and fun, spontaneous and voluntary activity, lacking extrinsic goals, being more an enjoyment of means than an effort aimed at a particular purpose, involving a participatory and active player. She also considered, however, another constitutive feature: contrast. Play keeps systematic connections with what is not play, which means it is related to some analogous non-play behavior that allows a contrast between what is play and what is not play. A behavior can only be identified as play when the one engaging in it can also adopt a similar or equivalent behavior in a functional or non-playful mode. By being able to perform either behaviors (the playful and the functional or non-playful), he or she can also understand the contrast between them.

Contrast recognition characterizes the play of wolf or dog puppies which quickly learn to distinguish between playful nibbling and real biting. Bateson (1972) suggested that these mammalian play patterns can only occur if the participating organisms are capable of reaching some degree of metacommunication, i.e., of exchanging signals conveying the message "this is play". According to Rivière and Español (2003), they reflect an *epiphany by denial*: the animal strips the hurtful bite of its effects, suspending it, and turning it into a playful and soft caress involving its teeth. Contrast recognition also characterizes the play of a 3-year-old child, who distinguishes between trotting on a broom used as a horse and riding a real horse (Rivière & Español, 2003). This kind of recognition appears to be a thread that phylogenetically and ontogenetically traverses play (Español & Pérez, 2015).

9.1.2 The Cradle of Play

Play is linked to know-how and to contact with peers. Infants play with what they know, and by playing, they unintentionally learn more. While play adopts increasingly complex and newer forms throughout ontogenetic development, the longitudinal observations we have made between the ages of 2 months and 3 years suggest that all those new forms are based on the first two types of play in human life: early

social play and sensorimotor play. Early social play is the first ludic activity involving others. It begins around the second month of life. It is pure social interaction, resulting from the adult's ability to improvise, brimming with musicality and involving the loving experience of being together. Sensorimotor play also emerges during the second month of life, it is the first play activity fully carried out by the baby through the growing mastery of his/her motor skills, it is pure corporeality, often occurring in solitude although it may also involve company.

Sensorimotor Play Halfway through the last century, Piaget (1946/2012) considered sensorimotor play to be the first form of infant play. This form of play begins at the end of the first month of life, when reflex adaptations give way to acquired habits. From that moment on, any infant behavior is likely to become play when repeated for pure assimilation, i.e., for simple functional pleasure. Play emerges when a motor scheme becomes independent of the purpose that provoked it: thus, the action of pulling the head backward to see something is detached from the interest of seeing something and is just performed for the pleasure of extending the torso and pulling backward. The motor scheme remains the same, although not motivated by the purpose of looking but by the bodily pleasure it produces. In the assimilatory activity infants find a unique source of pleasure, based on the mastery of a newly acquired skill (Piaget, 1946/2012). The movements that are guided by proprioceptive pleasure, combined with the emission of different sounds and noises, are the baby's first available play resources. The body is the first object of play.

Later, at about 4 months, the infant is able to coordinate his/her grip and vision patterns and can begin to grab and manipulate objects in the immediate environment. Once the baby can grasp, any novel object or event that is not perceived as threatening or dangerous attracts his/her attention, immediately becoming the object of his/her exploration. And after the object has been touched, probed, and tasted, the play activity may follow the exploratory activity. The motor schemes of grasping, throwing, or placing become independent of the purpose that provoked them in the first place, and are performed for the pure sensorimotor pleasure of repeating them over and over. At this point, the *pleasure of being a cause* is added to the functional pleasure of play. For Piagetian theory, any action scheme – piling one cube on top of another, knocking over a tower of cubes, folding and flattening a pillow – can be transformed into play as soon as it is mastered. In a protected play space, infants can try out their new skills and share them because adults often find their playful behavior funny, and join in.

Sensorimotor play enables the exercise and acquisition of action patterns and favors the development of complex cognitive functions. According to Piaget (1946/2012), it is the path to representation and symbol. In Bruner's (1964) terms, sensorimotor play permits the development of the most primary ways of representing past experience – enactive representations – that serve as a foundation for the development of iconic and symbolic representations. Exploration and sensorimotor ludic activity also favors the discovery of the potentialities of the physical environment, through direct perception and experimentation with object affordances (Gibson, 1969). According to the conceptual metaphor theory (Lakoff & Johnson,

1980, 1999), image-schemas are dynamic patterns that emerge as a result of the sensorimotor activity of our recurrent body movements in the surrounding space. For example, the VERTICALITY image-schema involves the recurrent experience of an upward and/or downward movement that links two spatial locations. The SOURCE-PATH-GOAL image-schema is an abstract structure resulting from the enaction of the purposeful movement originated in a spatial location that, after deploying a linear trajectory, reaches a final destination. Mandler (2005) and Johnson (2007) assume that image-schemas emerge during the infant's sensorymotor activity, such as ascending, descending, standing up, and so on. Likewise, based on non-linear dynamic systems theory, Smith (2013) considers that motor skills may be involved in a complex, multilevel chain of events through which a new structure or function is formed. Thus, rolling to a sitting position allows the child to hold objects and stabilize the head; he/she can then focus the attention on the object while holding and rotating it in his/her hands. By manipulating the object, the child can observe it from different spatial perspectives, which allows him/her to generate an abstract form and object substitution (characteristic of pretend play), as objects with similar abstract forms could be substituted. Moreover, sitting while manipulating objects seems to be the best condition for object name learning and verbal labeling (much more efficient than labeling an object held by the adult or close to the child but out of his/her hand). Also following non-linear dynamic systems theory, Bergen (2018) suggests that the development of sensorimotor play can be thought of as a dynamic system with the property of openness to exchanges with larger systems and the existence of control parameters that, under initial seemingly chaotic conditions, guide the development of self-organizing emergent patterns.

Early Social Play Around the second month of life, while sensorimotor play emerges, another relevant form of play becomes apparent: early social play. In the 1970s, Stern (1974), and Stern et al. (1977) explored the mutual co-regulation of the mother-infant dyad and discovered the purely playful character of some early interactions, whose goal was merely to stay socially involved. Early social play is an interactive play that takes place between adult and infant and is purely motivated by social interaction. It is a joyful play by nature, with social exchange and enjoyment as its sole purpose. Once initiated it brings all other external tasks to a halt while interpersonal focused events come to the forefront.

One of the first joyful activities adults offer their babies is playing with movement and sensations associated with it: they lift them up, cradle and toss them, moving their arms and legs around. In these situations of early social play, they also expose patterns of sound and vocal playing by clicking their tongues and lips, pairing these activities with tickles, rocking, and patting. Sounds, facial expressions, and movements are resources adults use to hold a baby's attention, generating expectation, driving the baby to different levels of arousal, some impossible to be achieved in solitude. Although parents are the initiators of these amusing activities, the baby also participates by answering and regulating the adult's actions. By directing their gaze toward the adult or looking away, they manage to influence those playing with them, regulating the amount of stimulation received (Garvey, 1977). In

the time period from roughly 2½ months to about 6 months of age, the infants' repertoire starts to include social vocalizations and facial expressions in response to parental behavior. During this time of full primary intersubjectivity development, the infant becomes a partner for social play (Stern, 1985, 2010).

The *repetition–variation* form often organizes early social play: parents repeat movements, sounds, words, or short sentences. This repetition is not mechanical or boring but varied in rhythm, intensity, kinetic or melodic contours; parents also modify the duration of pauses between repetitions, playing with their baby's expectations (Español, 2014; Papoušek, 1996; Stern, 1971, 1974, 1985). Play can take place in a varied array of ways. Sometimes, adult and infant engage in routinized interactions typified by unpredictable vocalizations, expectations, and facial expressions, as in a peek-a-boo game (Bruner et al., 1976). Other times, adults create suspense or expectation by moving an event in time back and forth, and multimodally emphasize the movement: extending the time in which the event occurs (tickling or catching the baby), they widen their facial expression and the opening of their eyebrows and eyelids; if they reduce the time, they shorten the distance between their own hands or between the infant's face and theirs (Stern, 1985). (Daniel Stern explained this in a roundtable discussion at the Philoctetes Center available at https://www.youtube.com/watch?v=-pTxptDPQzI&t=2879s). The adult frequently improvises infant-directed performances: he/she *invents* a motif and attunes with the baby to playfully model it. With each repetition, he/she embellishes or adds attractiveness using expressive resources shared with the temporal arts, such as Timing – the shortening and lengthening of a sound or movement. Very interesting things happen through infant-directed performances. Among others, the start of gestaltic experiences of image-schemas in inter-enactive contexts that will nurture later individual sensory-motor infant experiences. The infant's embodied experience linked to image-schemas shape specific features of early social cognition (for further details see Chap. 2).

Despite the fact that these modes of early social play have been recognized, there is relatively little research on the formats and structures of early adult-infant play (Fantasia et al., 2014). Likewise, these different modes of early social play are far from being exhaustive and exclusive categories. However, we can recognize a multimodal encounter between adult and baby in all these cases: the adult offers visual, sonorous, kinetic and tactile information and the baby responds and participates using different modalities. Their inherent multimodality turns these types of play into a privileged place for the expression and perception of vitality forms. As we will expose in detail in point 2, Stern (2010) described forms of vitality as dynamic feelings, an experience where a diversity of sensations coming from different modalities unite depending on their activation profile. Early social play is a ground where forms of vitality become fully evident. Sometimes early social play is almost purely a play with vitality forms (e.g., sudden almost explosive movements, accompanied by explosive laughter and vocalizations, when a mother tickles her child). Adults usually play with forms of vitality to prevent habituation and boredom. The repetition-variation form helps adults to level, modulate, and play off the infant's arousal (and theirs). To keep the baby in this optimal play zone, parents must adjust

their behavior; therefore, they have to modulate the forms of vitality performed. "The result is a theme-and-variation format of vitality forms" (Stern, 2010, p. 108).

As Stern (1985, 2010) notices, parents are a *sound-light show*, a spectacle to play upon their infant's state of arousal. As a stimulus becomes stronger, the arousal evoked is greater and the infant will try to turn away from it. If the intensity of the stimulus is too weak, the infant will not be sufficiently aroused and will remain uninterested and inattentive. The infant will be optimally excited to play and be happy in the moment just before the stimulus becomes too intense. Parents play with movement and sound modulating forms of vitality to enjoy intersubjective encounters with the baby. Feelings ranging from excitement to tranquility, from tension to relaxation, characteristic of early social play, are the same feelings that music and dance masterfully express. While the infant quickly becomes a partner for social play, there is an asymmetry that characterizes early social play: adults model the forms of vitality through the repetition-variation form, while the baby is an involved recipient who participates with general social behaviors – such as smiling, gurgling, vocalizing, and kicking – without which play would not exist. But the baby cannot yet modulate the forms of vitality (see Chap. 2). This will require going further along the path started in the cradle of play. We believe that early social play and sensorimotor play and their various combinations make up the cradle from which all play will later emerge. These early forms of play – with people, with oneself, and with objects – are the vital ground for subsequent play.

9.1.3 Pretend Play

Later in infancy, pretend play emerges, among others. This ludic activity has captured the attention of nearly all developmental psychologists for a simple reason: it links directly with symbolic ability, which is the axis of cognitive development. The controversy between Piaget and Vygotsky regarding the criteria for determining the genuine symbolic character of play is widely known (Kavanaugh, 2002; Español, 2004; Goncü & Gaskins, 2010). It can be synthesized as follows. From the Piagetian perspective, the precursor of pretend play or symbolic play lies in the playful execution of sensory-motor schemas that gradually unfold beyond their adaptive context to evoke an absent object through substitution, that is, by applying a schema that does not usually correspond to the present object but to the absent one being evoked (Piaget, 1956/2012). In contrast, according to the Vygotskian perspective, these first substitutions are just an appearance of substitutions and do not imply genuine symbolic substitutions. Symbolic substitutions only become possible toward the third year of life, once language emergence allows the construction of an imaginary scene where the child can play a social role (Vygotsky, 1933/1979; Elkonin, 1985). Beyond this classic debate, it is now agreed that presymbolic forms are linked to pretend play itself, and different forms of pretend play are linked to symbolic substitution. Symbolic play begins at the presymbolic or presentational level – known as *functional play* – when children begin to use objects out of their context (e.g., use

an empty spoon for eating) (Bates et al., 1979; Español, 2004; Garvey, 1977; Hobson, 1993; McCune & Agayoff, 2002; Rivière, 2003a, 2003b). With the irruption of substitutions, their combination and sequencing, pretend play emerges. Pretend play proceeds from simple, isolated substitutions (using a toy rake as a spoon) to multiple ones (feeding a toy horse from a plastic egg box) (Español, 2004; Garvey, 1977). The analysis of play that uses a synthesized version of Fillmore's case grammar – which focuses on the action and its cases – shows an orderly progression in case substitution, from the instrument case to the object, then to agent and recipient cases (Español, 2004). In tune with Bruner's (1990) seminal work on narrative thinking, this progression shows the child's pre-linguistic predisposition to organize the human activity experience in a narrative form. Role assumption articulated with language acquisition (playing teacher, mother, gardener) marks the appearance of role play (Elkonin, 1985). Functional play is generally recognized as emerging toward the end of the first year of life; from the eighteenth month on, when substitutions begin to appear in different action-scenarios, we are at the core of pretend play. It is worth noting that in all these recognized types of symbolic play, the focus is always on the figurative or representational content; there is always "something" being represented or figured. Although the child may play symbolically in solitude, the emergence and development of play is linked to social interaction and intersubjectivity skills (Göncü, 1993; Goncü & Gaskins, 2010), and is also culturally grounded (Kavanaugh, 2011).

In recent years, research on play has been oriented primarily to detect the sociocognitive skills to which it is related: language (McCune, 2010; Smith & Jones, 2011; Orr & Geva, 2015), Theory of Mind (Leslie, 1987, 2005; Meins et al., 2013; Karniol, 2016; Sung & Hsu, 2014; Lillard & Kavanaugh, 2014; Withebread & O'Sullivan, 2012), creativity and imagination (Sansanwal, 2014), divergent thinking and mathematical notions (Wallace & Russ, 2015). Other studies correlate it with the development of narrative ability (Guzdial, 2015), counterfactual and causal thinking (Buchsbaum et al., 2012) and executive functions (Thibodeau et al., 2016). It has also been linked to emotional development (Wieder, 2017) and emotional regulation (Hoffman & Russ, 2012). However, there has been criticism regarding the methodological and conceptual criteria used in these studies (Nicolopoulou & Ilgaz, 2013; Bergen, 2013). A review by Lillard et al. (2013) concludes that, apart from the functions of language, emotional regulation, and narrative capacity – where it may play a crucial causal role – pretend play is either one among many pathways to achieve the same developmental outcomes, or a mere epiphenomenon of the increasing complexity of other functions that truly drive cognitive developmental change.

A relative sense of stagnation seems to exist in our understanding of play in general and pretend play in particular (Bergen, 2013). Some point out that novel approaches to the study of play that go beyond conventional perspectives need to be developed (Fagen, 2011). We believe that the concept of forms of vitality – which locates the dynamic experience of movement as the original source of psychological life – may favor a novel approach to play.

9.2 Forms of Vitality

Vitality forms run through the above mentioned adult-infant early social play and they are also an important part of an interactive type of play typical of the child's third year of life, which we will present in point 3: forms of vitality play. In this section we will introduce the concept of forms of vitality, showing Stern's original proposal and summarizing research advances and current applications in several areas.

Stern attempted to define and describe the characteristics of the felt experience of vitality dynamics throughout his whole body of work in developmental psychology. He used different terms including *vitality affects*, *temporal feeling shapes*, and *vitality contours*, to account for those affective encounters between adults and infants that could not be described in terms of Darwinian emotions (anger, happiness, sadness, fear, disgust, surprise) (Stern, 1985, 2000). With the theoretical construct of forms of vitality, he reorganized and extended his ideas about the temporality and intensity of human dynamic experience to movement. "This (concept) adds movement, force, space, directionality and aliveness to the previous discussion of time and intensity" (Stern, 2010, p. 17).

Throughout his work, Stern recognized the influence of what Susan Langer (1967) called *forms of feeling* that are evoked by music (such as fading, exulting, easiness, rushing) and do not fall into the usual categories of emotion. In attempting to explain the feelings that music reveals, Langer brings in the words of the German composer and theorist Johann Hiller: "There are feelings … which are so constantly suppressed by the tumult of our passions, that they can reveal themselves but timidly, and are practically unknown to us. (…) Note, however, what response a certain kind of music evokes in our hearts: we are attentive, it is charming; it does not aim to arouse either sorrow or joy, pity or anger, and yet we are moved by it. We are so imperceptibly, so gently moved, that we do not know we are affected, or rather, that we can give no name to the affect" (Langer, 1954, p. 191). However, Stern does not use Langer's word to refer to the felt experience of vitality dynamics: "I am avoiding the term 'feeling' because of its close association with emotions" (Stern, 2010, p. 37) and proposes the term forms of vitality.

9.2.1 A Spontaneous Integration from Experiences of Movement

Forms of vitality is the last term proposed by Stern (2010) to capture how the human mind deals with dynamic experiences that are crucial in interpersonal encounters and time-based art experiences. He thought of the dynamic experience of movement as the original source of psychological life. Movement, and the proprioception experienced with it, are the earliest manifestations of being animate and provide a primary sense of aliveness. When describing what characterizes it, Stern came up

with the idea that vitality has forms. These forms emerge when the experiences of movement, and what he called its daughters – time, force, space and directionality/intentionality – come together all at once. A form of vitality is a Gestalt, a spontaneous integration emerging from experiences of movement and the elements that come with it. This fundamental dynamic pentad seems to be the way in which the mind grasps dynamic events. The perception of vitality forms is defined as the felt experience of force in movement with a temporal contour and a sense of being alive, of going somewhere. There is an evident primacy of movement in forms of vitality. Time and intensity become organized around the experience of movement.

As described in Chap. 2, there is some parallelism between Stern´s forms of vitality and the concept of *effort* proposed by Rudolph Laban in his system for movement analysis (Laban, 1971; Moore, 2009). Stern, however, conceived movement beyond kinetic dimension: forms of vitality encompass mental movement as well as physical action. He states that mental movement occurs when we think of something or feel emotions, since those experiences are not static. A thought can rush onto the mental stage and swell, or it can quietly just appear and then fade. Both physical and mental movement trace a journey, unfolding in time: "Mental movement, while it is happening, traces a profile of its rising and falling strength as it is contoured in time. This is its dynamic form of vitality" (Stern, 2010, p. 21).

Forms of vitality are modality non-specific. They do not belong to any particular content although they are generally associated with a specific one (e.g., emotions, physical or mental movements, actions, sounds, etc.). They differ from sensations, which are modality-specific (they originate from specific sense organs in the body), and from emotions, which are one of the possible contents to be shaped into the dynamic experience of a form of vitality. Forms of vitality can be grasped from experiences and be directly observed in the behavior of others. They concern the *how*, the manner in which things are done and not the actions themselves. Vitality forms characterize the style of an action and are detected on the basis of movement dynamics. We can bite, write or draw in an anxious, voracious way, or slowly and absent-mindedly. Every way of performing a movement or vocalization has specific arousal profiles. Forms of vitality are those arousal profiles grasped directly, without turning to inferences. Forms of vitality are, then, primary ways of relating to and understanding others; they are also basic modes of experiencing ourselves. Di Cesare et al. (2013) have suggested that besides the goal and the intention of the performing agent, vitality form is a third aspect an observer may capture when seeing an action done by another individual.

9.2.2 Forms of Vitality in Infancy and the Arts

Since his early work, Stern paid attention to the characteristics of the felt experience of vitality dynamics to account for those affective encounters between adults and infants that could not be described in terms of Darwinian emotions. Stern supports the idea that the ability to express and understand vitality forms is already present

in infancy. He described vitality forms as activation profiles, patterned temporal changes in intensity, sensations, and hedonic tone. And he portrayed the infant's interpersonal experience as a constant becoming of dynamic feelings difficult to put into words, which can only be described by means of dynamic terms like agitation, progressive dissipation, fleeting, and explosive, among others. Forms of vitality are amodal experiences where a diversity of sensations coming from different modalities unite depending on their activation profile (Stern, 2010). In this manner, if the mother says "all right, all right" while harmonically caressing her baby's cheek, the baby puts both actions together as long as they share duration, initial strength, and final abandonment. From both stimulations (tactile and auditory) the same vitality affect experience will arise.[1]

Forms of vitality are fused to a content; since they can be extracted from that content and transposed to another, they play a major role in situations of affect attunement. Affect attunement is a partial and purposely selective kind of imitation: dynamic features are faithfully imitated but using a different content, generally, in a different modality. In these matching behaviors the issue is not the open behavior but some temporal aspect of the baby´s behavior, such as rhythm, or some energetic aspect such as sound intensity, or some spatial aspect, such as a melodic or kinetic contour. This is one of Stern´s examples: a 9-month-old child is knocking a toy with his hand establishing a constant rhythm and the mother joins in the interaction by falling into that same rhythm saying "Caaa-bam"; the "bam" sound matches the child's knocking, and the "caaa" sound follows the preparatory actions of raising the arm and keeping it up in the air for an instant before the tapping occurs (Stern, 1985). Stern said that this type of matching behavior is what mothers do when they want to show their baby that they share what the infant is feeling. In 2010, as he reconceptualized vitality affects as forms of vitality, he stated that what is shared in affect attunement is a vitality form. The external behaviour being matched and the matching behaviour differ in form and modality, but are exchangeable as manifestations of a unique and recognizable internal state: the form of vitality. Affect attunement is thus a vitality form matching. It is based on the matching and sharing of dynamic forms of vitality across different modalities. Sharing another's vitality forms is probably the earliest, easiest direct path into the other's subjectivity. Also in 2010, through his concept of affect attunement, Stern linked his studies in early interaction with those –- at the time new – findings in communicative musicality (Malloch & Trevarthen, 2008) (see Chap. 2 for a description of this phenomenon). Stern argued that communicative musicality is largely based on the coupling of vitality dynamics between people. Being with another is accomplished by sharing vitality's dynamic flow.

From his first studies, Stern considered the baby's experience in early social play as resembling the adult's experience in non-figurative temporal arts (dance and music). Because forms of vitality are an experience where a diversity of sensations

[1] In Chaps. 5 and 6, the development of infant perceptual capacities linked to forms of vitality perception is addressed.

coming from different modalities unite depending on their activation profile, they require a multimodal context. Stern (2010) considered early multimodal social play as a privileged setting for experiencing vitality forms: the adult modulates for the infant the vitality forms of his or her movements, sounds, facial expressions, and the infant, when affected by them, responds in ways that often encourage the repetition-variation of the adult´s behavior (see Chap. 2). Thus the patterns of reciprocity that characterize second-person encounters gradually develop (see Chap. 1) and the infant gradually learns to move masterfully in the interpersonal world. More than once, he has pointed out that an infant perceiving their parent's behavior, a non-figurative dance spectator, and a music listener, are all in a similar position (Stern, 1985, 2000, 2004). According to Stern, adult-infant early social play and the temporal arts are the two contexts where forms of vitality are most clearly expressed and masterfully modeled. As we indicated in point 1.1, to regulate the infant's level of attention and arousal, and to generate expectation and surprise in early social play, typical of the first year of life, the adult plays with the vitality forms of his/her behavior: he/she modulates or varies the strength, direction or contours and speed of his/her movements and sounds. Also in the temporal arts, the dynamic profiles of sounds and movements (of speed, intensity, amplitude) create meaning in absence or beyond any figurative content (Stern, 2010). The infant's interpersonal experience, as well as the adult's experience with non-figurative temporal arts, is a constant becoming of dynamic feelings.

Stern (2010) explored the idea that time-based arts are largely about the dynamics of experience. The arts show vitality forms in a relatively purified manner: the dynamic features of a performance have usually been amplified, refined, and rehearsed repeatedly. Because vitality forms operate in all modalities and elicit similar felt states regardless of what modality they arise from, they are a privileged way in which arts collaborate with each other. Although all cultures have probably always used dance and music together, in the nineteenth century a new spirit emerged of exploring the combination of art forms to create new forms of art. Stern highlights the symbolist collaborations between poets, painters, and theater artists in that period and, a little later, the collaboration between the Russian film-maker Eisenstein and the composer Prokofiev or the collaborative work of the Bauhaus School involving painters, sculptors, architects, dramatists, film-makers, musicians, and perceptual psychologists. He also mentions the famous collaboration between the composer Cage, the choreographer Cunningham, and the painter Rauschenberg in the mid-twentieth century. When different forms of arts are juxtaposed, certain aspects do not translate well from one form to the next, but vitality forms are readily transferable between art forms because of their meta-modality and potential speed of modulation. This nature of vitality forms assures a common ability to render similar, but not identical, experiences. "The magic lies in pairing the similar with the not exactly the same" (Stern, 2010, p. 78). However, he considered each art had to identify its basic dynamics forms, and then invent codes for marking them.

Forms of Vitality in Music Stern (2010) describes some of the codes that music uses to mark dynamic forms and thus create vitality forms: the intensity (force), the

contouring of intensity in time, the stress or accent, among others. Using these dynamic markers, the rhythms and melodies with their variations and harmonies are imbued with the feeling of being alive and vital. The approach that understands music in light of the idea of vitality continued developing in recent years. Kim (2013), seeking an alternative to both cognitivist and emotivist approaches to musical expressiveness, considers expressive forms of music as forms of vitality, arguing that they are (co)shaped by musicians and listeners, giving rise to the aesthetic experience of music. Kim regards Stern´s forms of vitality as a key concept connecting musical experience with aesthetic empathy. In a more recent paper about the acquisition of instrumental musical skills from an embodied perspective (Kim, 2020), she considers that Langer´s theory of music and Stern´s concept of forms of vitality make it clear that pre-reflective forms of feelings have a structural analogy to the qualities of music. In this context, she argues that playing a musical instrument can become a means to understand the forms of our own feelings and to enhance our capacity for relating our own experience to the world and others. Shifres et al. (2012) specifically explore forms of vitality as the vehicle for styles of tango performance as music and dance. The study shows that the style of the orchestra when playing tango was transposed to the tango dancers. Shifres (2008) also studied forms of vitality, under the name vitality affects as they were still called at the time. He argued in favor of an understanding of music as a transmodal, intersubjective and second-person experience. The study utilized a fragment of Polansky´s *The pianist*, and was based on similarity judgments between musical auditory-only stimuli and audiovisual stimuli. They obtained the bewildering result that one of the musical performances (Olejniczak´s) was perceived as more similar to another one (Kissin´s) than to itself, because the image and the narration produced as a result a different experience of the same aural stimulus.

Forms of Vitality in Dance Stern (2010) highlights how the emergence of modern dance inspired a burst of exploration of new forms where dynamic experience was examined more and choreographers looked at non-linear, random, non-narrative form. In this way, the dynamic experience of movement became a focus of attention. He also refers to the problem of notation of vitality – the way in which movements are performed – and acknowledges that Rudolph Laban´s terms like effort or mood are closely related to the forms of vitality in movement. He stresses that to capture the uniqueness of an interpretation, Laban needed written dynamic indicators. In Chap. 2, it is possible to find a description of those indicators and their application in the movement analysis of the *dyadic dance* between adult and infant. Moreover, Stern´s ideas have undoubtedly had an impact on modern dance. In the early 1970s as Steve Paxton and his group were developing the technique of *Contact Improvisation*, he and Yvonne Rainer visited Stern´s lab (Stern, 1973). Paxton felt kinship with Stern´s slow motion study of gestures, sounds, and tone variations in the dialogues between infants and mothers. He realized that the work he himself was developing was very similar to what Stern had identified as affect attunement (Paxton explains this in the roundtable discussion at the Philotectetes Center, available online). In a recent conference, he recalled Stern and emphasized how important

those slow-motion videos of the 1970s had been to him, showing how fast we human beings are, much faster than we can perceive consciously. As a tribute to Stern, he came up with the expression *Stern moments* (Paxton, 2020). Also Sheets-Johnstone (2012) takes the idea of vitality affects, and links it with her phenomenological accounts of dance. Thurston and Slee (2017) use Stern´s conceptualization of vitality to create performances combining movement and poetry. They seek to activate vitality dynamics as compositional and performance principles, whilst exploring the gap concerning language-based arts, a problem Stern explicitly set aside. And art philosopher and critic John Rapko (2020) appeals to Stern´s conceptualization of vitality to think of Anne Teresa de Keersmaeker´s dance piece *Fase*. Rapko thinks that in this piece, the progression from synchronization through becoming out-of-phase and to re-synchronization of their movements towards the end, along with primitive expressions of vitality like clapping, jumping, and swinging of the arms, all that combined gives the piece its unusually intense sense of vitality.

Forms of Vitality in Theater Stern´s reflections about theater were mainly arrived at in relation to his long-standing collaboration with Robert Wilson, one of the most influential and innovative theater directors of the last few decades. *Bob´s breakfast: From in the mind to on the stage* (Wilson, Stern & Bruschweiler-Stern *apud* Stern, 2010) is a piece of dance-theater created as a result of an interview that Stern conducted with Wilson. In 2010, when analyzing forms of vitality in theater, Stern decided to focus on this play which consists mainly of movements on stage "to avoid the complication of a narrative line, and focus more purely on the vitality forms born of arousal profiles" (Stern, 2010, p. 89). However, in the last few years theater practitioners and theorists have found the concept of forms of vitality useful and rich for theater in much broader a sense than Stern himself. Wojciehowski (2014) applies Stern´s notion of vitality to analyze a scene in Shakespeare´s play *A winter´s tale*. She coins the term *vitality effects*, deliberate stagings of vitality forms intended to produce feelings and sensations in the viewer, which give the narrative much of its charge. She suggests that forms of vitality scaffold more complicated meanings, an idea akin to the observations of forms of vitality play in infancy as carried by some of the authors of this chapter in our first approaches to the phenomenon (Español et al., 2014, 2015). Along this line, Weeks (2013), and Nudler et al. (2018) analyzed scenes from Beckett´s *Krapp´s last tape,* comparing different productions. The latter created an ad-hoc observational code containing categories for forms of vitality in the actor´s movements, voice, scene sounds other than voice, and lighting. Stern underlined the extraordinary new dynamic forms of vitality in the lighting and scenic arrangements created by Robert Wilson. Based on Stern´s ideas on vitality, Bussières – a theater light designer – proposed to compose with light, and to consider a light score as a polyphony of temporally contoured events. Changes in light intensity are thus rehearsed, triggering emotional shifts and primitively establishing atmospheres (Bussières, 2018).

Forms of Vitality in Cinema Stern indicated that cinema has its own techniques to craft the dynamic feel of experience: the distance of the camera, through which arousal is regulated in a way akin to intensity (when the camera gets closer to the object, for instance, this is similar to a crescendo in music); the speed of alternation between shots in the editing process (if the speed is higher it would be equivalent to accelerating in music); the alternation between movement and stillness of the camera. He thinks that when these are combined with the vitality dynamics elicited by the other modalities in the film (movement, sound), the effects can be extraordinary. Also, the time of the vitality form and the narrative line can be placed contrapuntally. Recently, Barker (2019) appealed to forms of vitality and to the newer concept of forms of vitality play – the focus of this chapter – to flesh out Jean-Pierre Meunier´s classical theory of filmic identification.

9.2.3 Forms of Vitality in Neurological and Psychopathology Research, and in Psychotherapy

In the last few years, it has been argued that forms of vitality present specific neurobiological correlates. Di Cesare et al. (2013), using functional magnetic resonance imaging (fMRI), have identified brain areas underlying the recognition of vitality forms during the observation of actions done by others. They suggest that besides the goal and the intention of the performing agent, vitality form is a third aspect an observer may capture when seeing an action done by another individual. More specifically, the recognition of vitality forms appears to be linked to anatomical and functional areas independent of those described for the recognition of discrete emotions, such as anger and fear (Di Cesare et al., 2014). Di Cesare et al. (2020) indicated that the dorso-central insula is involved in both the expression and perception of vitality forms expressed through actions and speech. And that, unlike the neural circuitry involved in basic emotion recognition, the dorso-central insula has no subcortical connections with the hypothalamus, ventral tegmental area, ventral striatum, and nucleus of the amygdala. (It should be clarified that basic emotions are, unlike forms of vitality, discrete and carry a prototypical facial/body display and concomitant viscero-motor responses that do depend on the insula's subcortical connections with the hypothalamus, ventral tegmental area, ventral striatum, and nucleus of the amygdala.) These differences led Di Cesare et al. (2020) to support Stern's (2010) idea that vitality forms are not basic emotions. Thus, Stern's original psychological intuition that vitality forms are independent of discrete basic emotions seems to have found its neuroanatomical support.

The ASD (autism spectrum disorder) reveals the devastating effects that can result from the difficulty in interpreting or understanding the interpersonal world (Hobson & Lee, 1999; Rivière, 1998). There is evidence that children and youth with ASD experience significant challenges in identifying different forms of vitality. Rochat et al. (2013) found that autistic individuals made significantly more

errors than typically developing individuals on a task involving the recognition of the same form of vitality in different actions, or different forms of vitality in the same action. More recent research shows that children with ASD are nevertheless able to perform different actions modeled by different forms of vitality (Casartelli et al., 2020a). Strikingly, it appears that typically developing children present difficulties in identifying the forms of vitality expressed by children with ASD (Casartelli et al., 2020b). This last data brings novelty: it shows us that the difficulties in understanding vitality forms are reciprocal between individuals with ASD and typically developing individuals. In relation to psychotherapy, Stern questioned the reticence of traditional therapies like psychoanalysis to give room to the body, action, and movement, and their focus on the intra-psychic world in detriment of the inter-psychic. The cornerstone of his thinking was that body concepts provide the infrastructure for language, verbal concepts, and intersubjective communication. He derived this idea from the intersubjective world of the infant, and the perceptual characteristics of infancy. The lack of words, of language, permits and obliges the baby to remain immersed in the world of vitality forms for many months, bringing the opportunity to experience and learn a wide scope of relational procedures, of how it feels to be with others (Bruschweiler-Stern, 2017).

Stern stressed the urge to put vitality forms at the core of therapy as the elicitor of therapeutic dialogue (Boston Change Process Study Group, 2010). Intra-psychic events emergence can only be observed in the larger context of the inter-psychic, and inter-psychic happenings between therapist and patient need to be seen under the broader scope of intersubjective affiliation. When the therapist responds, he/she does a sort of affect attunement, so vitality affects are involved (Stern, 2010). Attunement provides a resonating response and a shared affect, therefore, it goes beyond empathy (Žvelc & Žvelc, 2008). During sessions, therapists attune to several different patient behaviors and experiences, providing the basic framework to process dissociated and unresolved experiences (Žvelc, 2012). According to Iacone, (2019) forms of vitality are a fecund construct in clinical settings that draw the therapist´s attention to the here and now, or what Stern (2004) called "now moments". A "now moment" is a short unit of time where something important is happening that will affect the future. It is an emergent property in a complex, dynamic system, a nonlinear leap in the process of the therapy session. A psychotherapy centered on forms of vitality is a highly integrative one, capable of embracing the multiple dimensions in which the therapeutic relationship unfolds and of linking the implicit to the explicit, according to the suggestions coming from the Boston Change Process Research Group (2010). When handled by the patient and therapist the implicit knowledge of each partner gets altered creating a new and different intersubjective context that changes the relationship. This process requires no interpretation and need not be made verbally explicit. It is part of the particular therapeutic approach that Stern has developed formulating the idea that *something more* than interpretation (Stern et al., 1998a, b) is necessary to bring about therapeutic change.

Forms of vitality have been enthusiastically picked up by non-verbal therapies. In the field of music therapy, Balducci (2020), who works with the Italian-school

psychodynamic-relational model (Postacchini et al., 2014), stated that Stern's theories have largely contributed to this model, particularly in the use of affect attunements that become technical elements for inter-subjective musical dialogue in professional practice. Stern´s writings became key for dance and movement therapists, laying down the foundational conceptual framework for their work (Vermes, 2011). For Lauffenburger (2020) it is dance movement therapy theory and embodied practice that enable the practical deploy of Stern´s *something more*, because of their placing the expressive moving body foremost in the therapeutic process. "We and our clients dance. No other creative arts, body-based, or verbal psychotherapy uses dance as its primary medium" (Lauffenburger, 2020, p. 7). Quoting Stern (2010), she suggests that dynamic experiences of space, time, force, and flow offer a physical-emotional life-affirming vitality which humans require to know our self and our place in the world. "These dynamic events are of course the essence of dance. Stern insisted that 'how' combined with 'feltness,' rather than story or logic, was key to development of a sense of self" (Lauffenburger, 2020, p. 8). According to Wengrower (2021), some experiences are nameless and they can be worked through nonverbally; the dance movement therapist can empathize and show understanding by reflecting or attuning to the movement: "(…) the recognition that feelings have forms and are not always speakable but can still be expressed and worked through is crucial" (Wengrower, 2021, p. 22).

9.3 Forms of Vitality Play

In recent years, the concept of forms of vitality has been incorporated into the analysis and interpretation of infant play. According to Trevarthen (2011, 2017), play is produced collaboratively and organized in time to create predictable experience *episodes*. Especially in physical "*rough-and-tumble" play* – where strong movements shared with the whole body provide immediate reward and feels exciting – children show and respond to Daniel Stern's vitality dynamics. Likewise, Bjørgen and Svendsen (2015) observed that vitality (energy communication) in children's physically active outdoor play has a contagious effect on the involvement and enjoyment of bodily active play. Pederson and Moe (2020) investigated how vitality forms are shared and modulated with others during preschool children play, finding that fluctuations in vitality forms provide a window to understand the ongoing reorganization of the intersubjective field between children. In these last two works, the possibility of mutually experiencing each other's forms of vitality is suggested as a way of sharing *present moments*. Stern (2004) coined the term "present moment" to refer to the time span where psychological processes group together very small perception units into global units (gestalts) that bring sense or meaning in the context of a relationship. Objectively, present moments last from 1 to 10 s, with an average of 3−4 s. Subjectively, we experience them as an uninterrupted now. They are the basic building blocks of relationship experiences. Modulating their movements' vitality forms with each other allow children to share present moments

(Pederson & Moe, 2020). In shared present moments, the conditions for physical and social reflectivity can be met, and experiences become intersubjective (Bjørgen & Svendsen, 2015).

Recently, some of us (Español et al., 2014, 2015) also observed infant play from the perspective of vitality forms. However, our focus was different: we concentrated on play activities where forms of vitality were at the center of the play – not just present as one more element. As recalled, Stern (2010) states that forms of vitality and the repetition-variation form are a meeting point between early social play and time-based arts. Therefore, we wanted to know if there could be later forms of play displaying these characteristics as well.

In point 1, we noted that those categories identifying the diverse modes of early social play are not exhaustive or exclusive. They are natural categories with diffuse boundaries and diverse examples thus some represent the category better than others (Rosch, 1978). Therefore, moderators are often cautiously placed next to the description of these types of play. For example, Stern (2010) points out that *sometimes* early social play is almost purely a play with vitality forms. Likewise, whenever there is an indication that the repetition-variation form characterizes early social play, an adverb is often added to emphasize it is not always the case. However, in Chap. 2, we introduced a type of early social play that always involves repetition-variation and the modulation of vitality forms: infant-directed improvised performances. These are short multimodal phrases spontaneously constructed by adults for their babies, based on the repetition-variation form of minor units or sound-kinetic motifs that are expressively modeled. The expressive and multimodal modeling in the motif´s repetition takes place in the spatial, energetic, and temporal dimensions. Infant-directed improvised performances are remarkably asymmetrical: the adult organizes the sounds and movements in the repetition-variation form and models the forms of vitality. The infant participates with general social behaviors (smiles, gurgles, vocalizations, kicks) encouraging the adult to generate the performances, but is not involved in the play construction. For example, a casual movement such as removing a blanket covering the baby (the motif) is twice repeated in a variety of ways. In a repetition, the adult adds staccato vocalizations composing a contrasting unit along with the movement of sliding the blanket. Chapter 2 presents a microanalysis of this and other infant-directed improvised performances by an adult-infant dyad in a local urban setting. Chapter 5 reviews an infant-directed performance by an Afro-Colombian dyad from the Guapi community, where while facing his baby, a father lifts, holds, and places her on his lap. This sequence will then become the motif to be repeated with variations in the quality of the movements and vocalizations. Infant-directed improvised performances develop during the first year of the infant's life, reaching their peak around mid-first year.

Towards the third year of life, the child becomes an active participant in different kinds of play. In earlier works, we decided to observe whether, during this developmental period, play interactions where the child actively participated in the organization of the repetition-variation form and in the modeling of vitality forms emerged. Reddy (2008) proposes a second-person methodology for the psychological research of intersubjective phenomena (as opposed to the distanced and objective

third-person studies, and the subjective studies in first person). She suggests that the knowledge we can generate on the phenomena of interpersonal involvement essentially depends on the intersubjective relationship established, i.e., the person-to-person genuine encounter. In our play studies, we followed Reddy's psychology of intersubjective phenomena's methodology proposal. In all of them, we focused on the researcher-child dyad (a particular case of the classic adult-child dyad). When studying play from a second-person methodology, we found that throughout the third year of life, in the context of a flowing interpersonal interaction and social engagement, there is an emergence of playful exchanges that preserve the common features of early social play -particularly infant-directed improvised performances- and temporal arts.

9.3.1 Towards a Categorization of Play with Forms of Vitality

In Español et al. (2014) we applied the *constant comparative method* (Glaser & Strauss, 1967) to observational records of interactive adult-child sessions, generating revisable and open categories to detect the properties of these ludic interactions, where the child actively participates in the organization of the repetition-variation form and in the modeling of vitality forms. We studied the period comprehended between months 24 and 36, using a *convergence approach* (Bell, 1953), involving the combination of cross-sectional and longitudinal techniques, overlapping age periods to accelerate the longitudinal study (Duncan & Duncan, 2012). Two girls – Coty and Kiara – transiting their third year, participated in the study. Coty was 2 years old in the first session and 2 years, 5 months old in the last one; Kiara was 2 years, 4 months old in the first session and 2 years, 10 months old in the last one. One of the girls is an only child; the other has two older siblings. They were both born to Argentinean parents. The adult was the first author of this chapter (S). She was particularly attentive to promote situations of play involving sound and movement as well as figurative play. For that reason, she took special care to ensure that child and adult had plenty of space to move around. They frequently stayed on the floor.

The characteristic aspect of this second-person procedure is an active participation of the researcher, generating a special kind of interaction that differs from everyday interactions. Choosing a researcher-child dyad (a particular case of the classic adult-child dyad) permits to simultaneously take two fundamental perspectives on knowledge generation in reference to interactive systems: *emic* and *ethic* – correspondingly, internal and external to the analyzed system (Riba, 1990). The researcher can register her own experiences and reactions on-line, generating changes around them, and thanks to the availability of registered audiovisual resources she can later observe them, retrieving her own previous experience. A set of appropriate objects for play activities was displayed in front of the girls: objects of conventional use, dolls that could easily have roles assigned to them, and objects of non-conventional use.

The interactive sessions were recorded at the girls´ homes every 15 days, summing up a total of 28 sessions with 27, 44 filmed hours. The adult started each session by offering toys and inviting the girls to play. The toys used and the activities displayed depended on each girl's spontaneity and motivation. However, the researcher encouraged play activities by initiating and inducing the girls to participate. Each session lasted 40 min. For further details on the procedure and to access the complete microanalysis of the scene, see Español et al. (2014).

The iterative analysis of the recorded footage allowed us to create a general category, divided into two groups and disaggregated into four subordinate categories. The general category forms of vitality play (hereafter FoVP) includes every playful interaction where the girls actively participated in the organization of the repetition-variation form and in the modeling of vitality forms. These interactions differ from interactive physical and locomotor play: although forms of vitality are present in the latter, they do not involve the pleasant exercise of forms of vitality repetition variation co-regulation that generates a present, reciprocal, bodily, and aesthetic interpersonal encounter. The iterative review of the general category showed that sometimes FoVP appears in a simple way and sometimes combined with figurative or symbolic content. Consequently, the category was divided into two groups: simple FoVP and combined FoVP. Simple FoVP arises spontaneously, achieving a high level of organization at the time of execution. The whole activity unfolds around the varied repetition of sounds and movements, which becomes the unique focus and essence of this play, in detriment of any figurative or symbolic content. Combined FoVP also emerges spontaneously, but involves the development of a narrative *theme* and some fictional aspects. In these cases even though sonic and kinetic forms of vitality will constitute the focus of attention, they accompany the figurative or symbolic content. The iterative observation of simple FoVP revealed that FoVP sometimes relies on a pure real-time improvisation of sounds and movements, and sometimes on pre-established organized forms of sound and movement. Improvised FoVP, similarly to what happens in the infant-directed improvised performances presented in Chap. 2, consists of fleeting events that emerge, unfold, and get organized just once, to disappear right after. They can be brief or long, they can be organized in a very simple or extremely complex fashion, and the *making together* may involve similar or dissimilar participation. Ritualized FoVP are created out of pre-established organized forms of sound and movement. They include highly ritualized types of play, such as action songs (described in Chap. 3), nursery rhymes, advertising runs, choreographed exhibitions, and all kinds of events where song and dance modes belonging to the children's cultural dance and music repertoire can be recognized. The iterative observation of combined FoVP revealed that FoVP combines with the two traditional forms of symbolic play: pretend play and role play. Consequently, two subcategories were generated: pretend play & FoVP, and role play & FoVP. Figure 9.1 shows the general FoVP category, the two groups into which it is divided, and the four disaggregated categories.

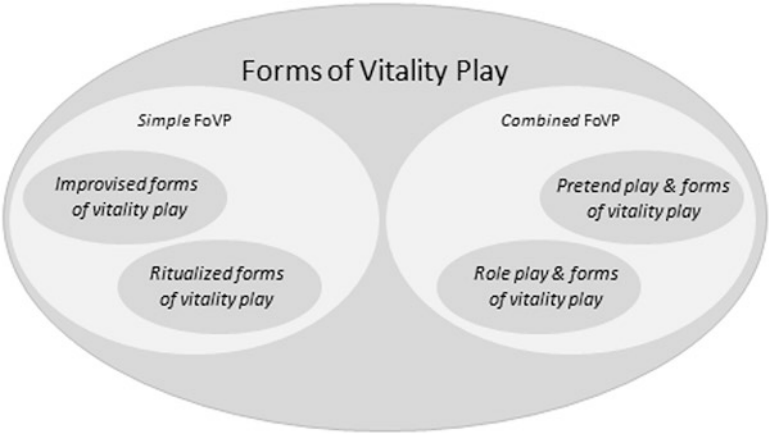


Fig. 9.1 The general FoVP category, the two groups (simple FoVP and combined FoVP), and the four subcategories (improvised FoVP, ritualized FoVP, pretend play & FoVP, and role play & FoVP)

Operational Definitions of the General Category and the Four Subcategories

FoVP: It is a pleasant and joyful, either soft or abrupt, vital and sensitive kind of play where an intimate and vivid sharing experience takes place between those who play it. In FoVP, adult and infant elaborate units of movement and/or sound according to a repetition-variation form. The units (or motifs) are repeated at least twice. The variations of the sound-kinetic motifs are carried out in the spatial, energetic, and temporal dimensions of sound and movement. These are the same dimensions where the motives of the infant-direct improvised performances vary. We summarize here the description of the dimensions given in Chap. 2 (more detailed information can be found there).

Spatial modeling includes:

- For sound: Melodic contours and timbre variations.
- For movement: Shape variations, as described in the Laban system of movement analysis: rising-descending, spreading-enclosing, and advancing-retiring.

Energy modeling comprises:

- For sound: Degrees of relative sound intensity (Pianissimo [very soft], Piano [soft], Mezzopiano [medium soft], Mezzoforte [medium loud], Forte [loud], Fortissimo [very loud]).
- For movement: Effort variation, as described in the Laban system of movement analysis (1) pressing (direct – sustained – strong), (2) gliding (direct – sustained – light), (3) punching (direct – sudden – strong), (4) dabbing (direct – sudden – light), (5) wringing (flexible – sustained – strong), (6) floating (flexible – sustained – light), (7) flicking (flexible – sudden – light), (8) slashing

(flexible – sudden – strong). Each of the eight basic types of effort can be performed in a continuum flow (the fourth factor): from a free flow (difficult to stop) to a bound flow (easy to stop).

Temporal modeling comprises:

– For sound and movement:

Speed variations: *Ritardando, accelerando, rubato*.
Tempo variations: Largo, grave, lento, adagio, andante, moderato, allegro, vivace, presto, prestíssimo.
Behavior organization according to temporal amodal information: Duration, rate [chronometric density], and rhythm.

In forms of vitality play, a brief sound/kinetic motif is repeated with variations in the spatial, energetic, and/or temporal dimensions. Figure 9.2 presents a diagram of the dimensions where the repetitions of the sound/kinetic motifs of FoVP vary.

Improvised FoVP: It is an improvised frame where units of movement and/or sound (motifs) are repeated varying the spatial, energetic, or temporal dimensions of sound and/or movement. The whole activity unfolds around the varied repetition of sounds and movements, making it the unique focus and essence of this play, in detriment of any figurative or symbolic content.

Ritualized FoVP: It is a frame where rhythmic patterns, forms, and dynamics of recurring movements and/or melodic contours are organized around a cultural established song or dance. The whole activity unfolds around the varied repetition of sounds and movements, making it the unique focus and essence of this play, in detriment of any figurative content.

Pretend Play & FoVP: It is a frame where at least one case of substitution occurs, together with repeated units of movement and/or sound (motifs) varying. The activity evolves around the varied repetition of a sound or/and movement motif, making it the core of this play but mixed with fictional actions.

Role Play & FoVP: It is a frame where a simulation of a social role occurs together with repeated units of movement and/or sound (motifs) varying the spatial, energetic, or temporal dimension of sound and/or movement. The activity evolves around the varied repetition of sounds and movements, making it the core of this play but mixed with fictional characters or role play.

Illustration of the Four Subcategories

Since the four subcategories of play may not be easily imagined at first, they are illustrated with observational accounts taken from Español et al. (2014).

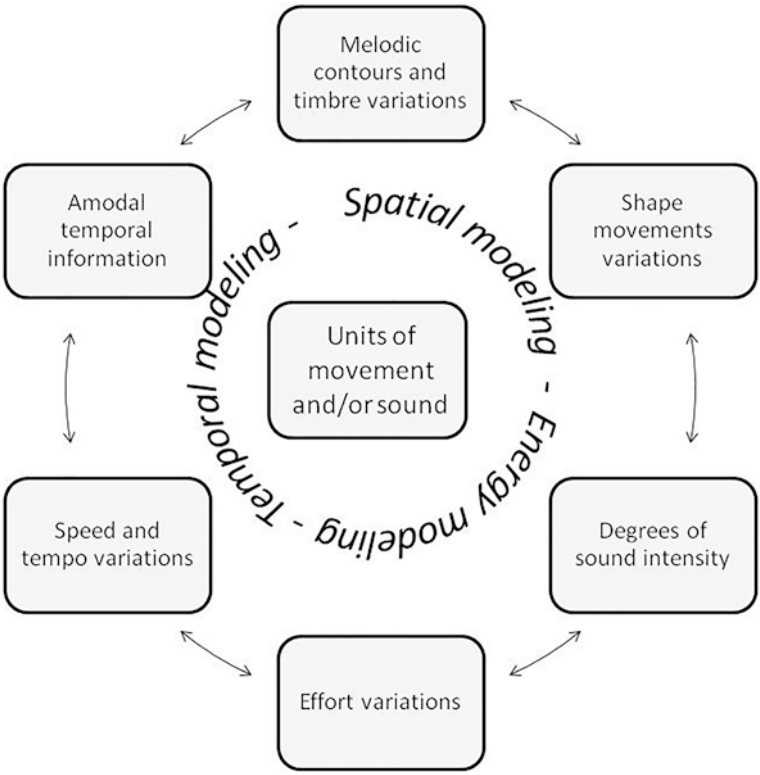


Fig. 9.2 The three dimensions of the modeling of sound and movement units. The double arrows indicate the absence of a prefigured order or hierarchy in the selection of the components of the dimensions

Improvised Forms of Vitality Play

Example 1. Coty, at 2 years, 3 months, slowly and sinuously lets her back roll down a couch toward the floor, and then quickly straightens herself back up. She repeats this sequence three times, varying the form and duration of her movements, and the vocalizations that accompany every repetition. The adult performs similar movements and melodic contours.

Example 2. When Coty is 2 years old, she is sitting on the ground approximately 1 m away from the adult and, for 40 s in which they smile and look at each other's face most of the time, the following play with two springs unfolds: The child and the adult hold the springs' tips in each hand and shake them, generating a wave-like movement in the horizontal plane. The girl gets distracted by saying something but keeps going until she freezes the movement. The adult also stops her movement vocalizing "Up up". Suddenly, the girl begins to move the spring quickly, and exclaims "Aaaaaaaaa". The adult follows her in unison, imitating the girl's

movements and vocalizations. Coty abruptly stops her movement vocalizing “Ia”; the adult follows, exaggerating the mouth opening expression. In this way, the play’s motif is composed by a shaking of the spring, a vocalization, and a pause. The adult repeats the motif again by shaking the spring, varying the vocalization (now says “Traaaaaaa”), and stopping the movement abruptly. The girl does the same thing, in silence and with a wide smile. They repeat it again, but now the girl pauses a moment after the adult, who softly vocalizes “Ia”. Then she suddenly lets the spring fall and looks at the adult, who imitates her. Immediately after, the girl lifts herself up from the floor and points toward other toys.

Example 3. Kiara at 2 years and 4 months, together with the adult, takes turns to draw while vocalizing invented words for approximately 2 min. She draws some lines and then asks the adult to draw a girl. The adult playfully refuses and says she liked something Kiara just did (in reference to the drawn lines) and proceeds to draw some lines while vocalizing “Ta, ta, ta, ta, ta”. The girl responds making some new lines on the paper and vocalizing “Ta, ta, ta, ti, no, ve, de”. The researcher gets back at her by doing the same thing and says “Ta, ta, ta, pi, no, ver, de”, she laughs (since “pino verde” is the first non-invented vocal expression in the play, meaning/green pine tree/). The girl makes a slightly different movement and adds a different voicing, increasing the tone and movement’s intensity but maintaining the rhythm and melody of the motif. The adult does the same thing, and they smile at each other. Then the child draws the lines silently and when she is done she exclaims: “¡mirá!” (Look!). The adult draws the lines and vocalizes “Ta, ta, ta, ta, ta, ta, ta”; but the girl corrects her and says: “Es así’ (It goes like this) and goes back to drawing silently. The adult redraws the lines while vocalizing; the girl corrects her once again, until the adult draws silently during her turn. Then she changes the paper, draws a circle, and re-establishes the previous motif by making stripes while vocalizing “Ta, ta, ta, ta”. The girl, smiling, draws while exclaiming “Ta, ta, ta, ta”, strongly raising the sound and movement’s intensity; the pitch sound also raises up to a point where the adult cannot imitate the girl (her vocal registry does not get that high) and lowers the tone. Then, they imitate each other twice until the adult abruptly changes the volume of her voice turning it very soft. The girl loudly replicates with a “t, t, t” sound. They keep going for many brief repetition-variation cycles where they play with the dynamics of vocalizations and movements: the adult says “Tu, tu, tu, tu” performing short traces at the same pace; the girl does the same thing. Then the adult with a legato, soft, and ample movement, draws a circle saying “Uuuuuuuh” with the same duration and characteristics. The girl doesn’t follow this proposal and generates a new motif: with two fast and sturdy traces, she draws a cross while vocalizing “Mi, chá”. The adult reproduces this vocalization and movement. The girl repeats the vocalization varying it by saying “Tan, tiá”, the adult repeats the same thing. The girl varies it again by saying “Santiao”, the adult responds “San, tia, go” (second non-invented verbal expression, Santiago is a male name) lastly, the girl exclaims “San, tia, jj” and the pencil tip breaks while she is performing some rough traces. They both laugh. The play is over.

Ritualized Forms of Vitality Play

Example 4. Coty at 2 years, 1 month picks up a pile of rectangular-shaped wooden pieces held together with a ribbon (object with no clear conventional use) and sings "Un tallarín, un tallarín" (a spaghetti, a spaghetti) flapping her arms. When she finishes singing this phrase she looks at the adult who asks, smiling: "¿Un tallarín?" (a spaghetti?). The girl goes back to singing "Un tallarín, un tallarín", while rotating her arms and torso to the sides. And continues singing "Que se mueve por acá" (moving over here), moving the toy with shorter movements in front of her visual field, accompanying the vocalization *rallentando* "Por acá" (over here), ending with a sudden lowering movement that hits the object on the table. The noise made by the impact is synchronic with her last vocalization. Then she freezes her movement, stares at the adult saying "ap" and the adult responds "ap". The girl starts singing a popular children song. The adult asks "¿Un tallarín?" assuming that the girl is representing the spaghetti with the pile of rectangles. The girl does not answer: she's not pretend-playing. She is busy playing with her own sounds and movements, trying to make the dynamics and intensity match her sounds and jointly vary on a preestablished performance plan.

Example 5. Kiara at 2 years and 5 months, sings *Tea Song* – from songwriter María Elena Walsh. The little girl, who was pretend-playing with some dolls, adds music to the dolls' stroll by humming an invented song. This constant humming drifts on the *Tea Song* melody and the girl begins to sing the song's lyrics (although mixing parts of them) while performing several actions (moving the puppets, placing some ribbons on the table). Suddenly, she remains still in front of the adult (who is sitting on a little chair), she looks at her and sings adding some gestures in accordance with the song's text. For example, when she says "La nariz dentro de la taza/eso no se ve/ yo no sé por qué" (the nose inside the cup/you can't see that/I don't know why), she touches her own nose and shakes her head no. She repeats this part of the song three times. In the first repetition the adult sings with Kiara, but then she silently pauses while staring and listening to the girl. In the closing of the last repetition, the adult starts singing with the girl again. The adult continues the song and the girl sings along in a slightly delayed manner since she does not remember the lyrics very well. At the same time, the girl begins to play with a few cups; the adult carries on and they both repeat five times the song's final phrase ('yo no sé por qué') (I don't know why) while tapping the glasses on the table at the time they say "pum". The hit and the "pum" sound are performed in unison while maintaining the phrases' metric. The adult makes a *rallentando* that indicates the song's final closure. They remain in silence and start a new play. The entire song lasts approximately 1 min, beginning when the girl starts singing and ending when the adult closes.

Pretend Play & FoVP

Example 6. At 2 years and 1 month, Coty grabs and moves the spring like a steering wheel saying "El paseo de papá" (daddy's ride). When the hand hits the spring twice she says "Toco bocina pip pip"' (I play the horn pip pip); repeating "Pip pip" and hits it again. The girl repeatedly alternates her gaze between the adult and the spring. She begins to rotate it with sinuous arms and trunk movements singing "El paseo de papá", then she tiptoes quickly and repeats "El paseo de papá, el paseo de papá". She keeps running to the door and stands in front of it. The adult replies by singing "El paseo de papá". At that point the girl leans the spring on the door frame and moves it upwards by stretching herself to finally move it downwards. The spring symbolically replaces the wheel and the sinuous movements of arms and trunk seem to evoke the path of a car driving on the street. The repetition-variation of sounds and movements made by the child beautifies and enriches her pretend play.

Example 7. At the same age, Coty grabs a wooden doll and says "A bailar nos vamos" (dancing we go), she shakes it up and down, hitting the table with it more than ten times, accompanying these movements with the vocalization "La la la". Toward the end, she accelerates the beating pace and changes her vocalization to "Uh ha ha" further accelerating the hitting, until she stops. The doll falls, Coty says "Upa". Then she lays it, face down on the table, and says "A dormir" (go to sleep), and taps its back twice with her hand synchronously saying "Sh sh". She repeats the motif giving three strokes adding "Sh sh sh", she looks at the adult who is also doing "Sh sh sh" and turns to repeat it in the same way. She rolls the doll back up, lifts it upwards and says "A bailar vamos" (dancing we go) once again. The girl attributed animated beings' traits (sleeping, dancing) to a wooden doll and got the doll to do animated actions by repeating and varying sound units ("la la la", "sh sh sh", etc.) and movements performed on the doll (bob up and down, pats on the back, etc.).

Role Play & FoVP

Example 8. Kiara at 2 years, 9 months is playing with some wooden dolls (two witches and two robots), and says that the two witches are the researcher and the camerawoman, while she and her sister are the robots. Then she approaches her sister Milena (who was watching in a non-participatory way) and starts singing with a nasal, strained voice "Somos las robóticas" (we are the robotics) repeating it eight times in a variety of ways, dancing, moving the dolls and looking at them, making them dance. The phrase's melody is invented by the child. Then, while still singing, she takes two witch dolls and adds "Ustedes son las brujas" (you are the witches) and strongly hits them twice against the table. And keeps on singing "Las brujas no existen más/las brujas no bailan más/las brujas no bailan más" (no more witches/witches do not dance anymore/witches do not dance anymore). The researcher and the camerawoman constantly stare and smile at the girl. Then she grabs the doll representing her sister and says "La voy a peinar a Milena" (I will comb Milena's

hair), she runs her hand through her hair and then leaves the room. The sequence lasts about a minute and a half, with many repetitions of various sound motifs or units ("somos las robóticas", "las brujas no existen más") and also kinetic (movements that accompany these verses). All sung phrases are created on-line by the girl, and her expressive features (sound and movement) characterize the characters (witches and robots). The researcher, the camerawoman, and the sister are there, smiling and accompanying her, while she displays and gives life to the whole scene with her own resources.

Comments on the Illustrations of the Four Sub-categories

In Examples 1, 2, and 3 there is an evident adult-infant joint construction – lacking a verbal mediation – of the motif, which is repeated with variations. They all involve the modulation of sounds and movements through the varied repetition of short motifs. However, they differ in their duration and complexity. Examples 1 and 2 are relatively short (less than a minute long). Instead, Example 3 is a complex and extensive play (over two minutes long) formed by many motifs and varied repetitions. Its peculiarity resides in the creation of shift alternation cycles, non-figurative drawings, and the inclusion of invented words in the constitution of motifs and variations. Child and adult establish mutual imitation cycles of sounds and movements with turn taking, introducing variations in the movement's dynamics, intensity, and form, and in the intensity and articulation of the sound. The micro-analytical study of vocalizations revealed that girl and adult not only shared the time pattern of their performances, but toward the end of their moment of playing, they achieved an organization of their performances in a shared metric hierarchy (to view a full analysis see Bordoni & Martínez, 2011). In the movement microanalysis of a fragment of this work, we saw that the movement´s shift alternations were organized by the adult's movement-stillness pattern and by the girl´s back and forth movements (Bordoni and Español, 2011).

In Examples 4 and 5, the girls sing nursery rhymes typical of their local cultural repertoire. Responding to the song's structure, they perform a personal version adding the described movements. This type of play illustrates how the progressive constitution of dialogical forms shared by the partners results in a clear open behavior that can be appreciated by other people, by integrating ritualized forms of exchange and local cultural elements. Examples 1, 2, 3, 4, and 5 are simple FoVP events. They show the independence of this type of play where the activities elaborated lack figurative or symbolic content. On the contrary, Examples 6, 7, and 8 show that FoVP mixes with the child´s symbolic activities generating combined forms of play. The child uses FoVP as a resource to elaborate symbolic events. The varied repetition of movements and sounds merges with figurative and symbolic content, enriching Examples 6 and 7 with symbolic substitution. In observation 8, the varied repetition of movements and sounds helps to distinguish characters or social roles. In both cases, FoVP beautifies and enriches fiction development. These types of play also differ in their duration and complexity. Some of them last more than 2 min,

achieving a vast compositional complexity. Like in Example 8, where the sung phrases improvised by the girl present clear melodic, rhythmic, timbre, and textual differences according to the characters of the play – the robotic girls and the witches. In this play, the quality and form of the movements accompanying the song and the repeated substitutions in the agent of action (see Sect. 9.1.3) also indicate that the compositional complexity is almost entirely in the child's hands. In contrast, other play activities last less than a minute exhibiting a simpler composition, as in example. 3.

The Rate of Forms of Vitality Play

As we illustrated above, children go beyond early social play and develop the ability to actively manipulate or modulate forms of vitality in their dialogical exchanges. Children go from perceiving to purposefully manipulating forms of vitality. The activity´s frequency can be used to identify what the child is dealing with, giving valuable clues concerning this activity in that particular moment. The number of activities and their frequency can be used to obtain information about changes in the development of exploratory and play activity in children (Vlietstra, 1978; Rivière, 2003b). Identifying the frequency of FoVP and contrasting it with that of symbolic play – traditionally the most studied play in developmental psychology – could help us ponder its relevance for child development.

To address this question, Español et al. (2015) conducted a longitudinal study using the rates of FoVP and symbolic play during the third year of life as indicators. They calculated the corresponding rate of play events per filmed hour (Σ play events/filmed hours) for each quarter in the third year of life. First, the rates of the categories simple FoVP, combined FoVP, and symbolic play were analyzed. The simple FoVP category includes improvised-FoVP and ritualized-FoVP, the combined FoVP category includes pretend play & FoVP and role play & FoVP. The symbolic play category was defined as every ludic activity where at least one substitution or simulation of some social role is present. Table 9.1 shows the quarterly rate of each of these categories.

The three types of play studied are present from the beginning of the third year of life. Table 9.1 shows that symbolic play holds the highest rate in comparison to simple FoVP and combined FoVP. During the second quarter, the rate of symbolic play decreases to fall below the rate of simple FoVP. In the third and fourth quarters, symbolic play events notoriously increase. A constant rate of FoVP is present from the beginning and throughout the third year. In the first quarter of the third year, this rate value is lower than symbolic play´s, while in the second quarter FoVP's rate outgrows that of symbolic play. Later, in the third quarter, the relationship shifts back to the one in the first quarter; and finally, the rate of symbolic play is twice and a half FoVP's rate. Combined patterns of play show lower rates in comparison to the other categories, and a U turn trajectory during the studied period with the lowest rate. During the third year, simple FoVP´s rate remains constant. Its rate value is similar to that of symbolic play during the first three quarters, but is duplicated by

the symbolic play's rate in the last quarter, precisely during the period where symbolic play is the most frequent.

Subsequently, the rate of the four subcategories was analyzed: improvised FoVP, ritualized FoVP, pretend & FoVP, and role & FoVP. Table 9.2 presents the quarterly rates of the four subcategories.

The rates of improvised FoVP and ritualized FoVP follow different trajectories during the third year of life. Even when improvised FoVP's rate is higher than that of ritualized FoVP, they evolve inversely during the third year: while improvised FoVP decreases in the second half (third and fourth quarter), ritualized FoVP increases. The rate of pretend play & FoVP remains relatively steady, always below the simple form's rate. The combined form of role play & FoVP appears later, in the third quarter of the third year with a tendency to increase toward the fourth quarter.

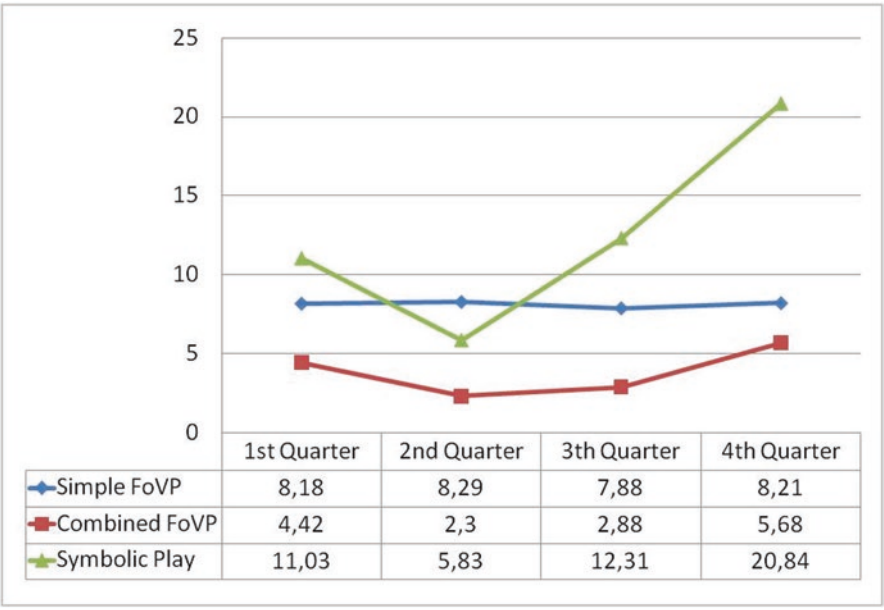
9.3.2 Comments on Rate Trajectories

The rate trajectory of FoVP reveals that elaborated activities developing independently of any figurative or symbolic content are not sporadic or unusual, and do not tend to disappear during the studied period. On the contrary, the elaboration of movements and sounds in a repetition-variation structure (independently of any figurative contents) is an activity that takes a good proportion of infant play and remains constant during the third year of life. Its constant rate throughout the third year suggests it is a companion to development.

Improvised FoVPs are usually fleeting, ephemeral moments. Generally, they reach a substantial and beautiful form, but right away they disappear and are not re-established later. In the few cases in which they stay from session to session, joy of recognition runs along; but in the end, they all fade away. The high rate of improvised FoVP suggests the relevance of unplanned, spontaneous, and continuous performances of sound and movement in the co-construction of the researcher-child interaction. Improvised FoVP decreases in the second half of the third year of life, while ritualized FoVP increases. Arguably, the change in the trajectory of the ritualized FoVP rate reflects the girls' appropriation of culturally decanted products. Conventional forms of singing and dancing are *culture packages* for child development. When ritualized and culturally established songs and movements are introduced, there is an evident tendency in the exchanges to continuingly construct shared meaning and public cultural activities of meaning-making. In this sense, we can think of ritualized FoVP as a road toward establishing cultural shared activities. The social value of these decanted products of culture undoubtedly favors their execution and conservation in the child's repertoire.

The relationship between pretend play and pretend & FoVP suggests that children use FoVP as a resource to elaborate symbolic events when playing pretence, and that once mastered, pretend play tends to be performed more frequently unaccompanied by FoVP. The studied period allowed us to witness the particular temporal relation between a simple form of play and a combined one. Few cases of role

Table 9.1 Rate of simple FoVP, and symbolic play



	1st Quarter	2nd Quarter	3th Quarter	4th Quarter
Simple FoVP	8,18	8,29	7,88	8,21
Combined FoVP	4,42	2,3	2,88	5,68
Symbolic Play	11,03	5,83	12,31	20,84

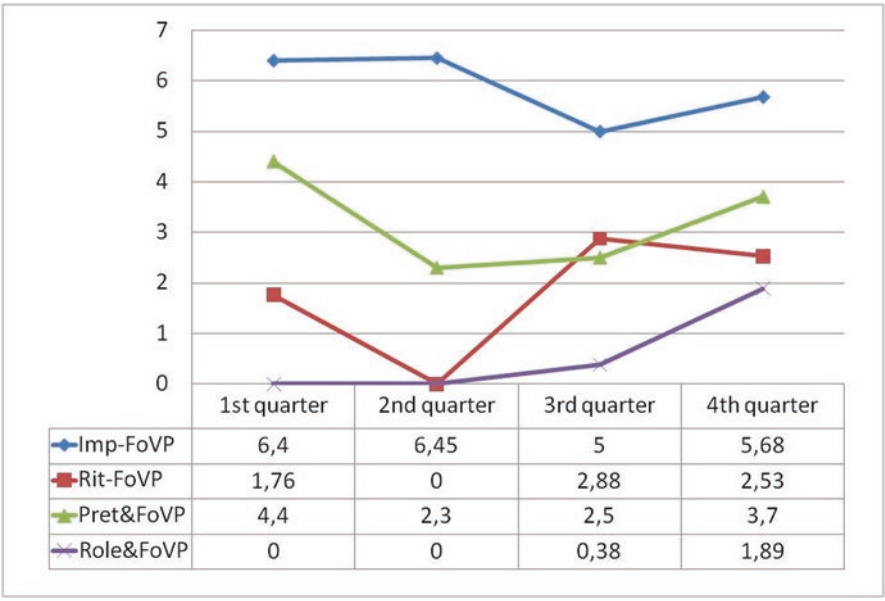
Quarterly rate (Σ play events/filmed hours) of simple FoVP, and symbolic play during the third year

play & FoVP were observed; all of them toward the end of the third year. The emergence of role play & FoVP during the 3rd quarter might be indicating that play develops by means of a combination between previously established types of play (FoVP) and novel ones (role play).

New skill development resulting from the combination of already acquired skills was observed in ontogenetic phenomena preceding the described game. For example, the communicative pointing gesture is a Gestalt comprising the pointing of the index finger while alternately looking at the event being pointed out and at the person to whom the gesture is directed. Gaze alternation – which indicates joint attention – appears several months before the pointing gesture. Pointing without alternation or exploratory pointing in which the index finger touches the object also occurrs a few months earlier. Once gaze alternation is established, this previous pointing can be incorporated into the structure of gaze alternation, leading to the transition from joint attention to communication through gestures (Español, 2004).

FoVP is present from the third year of life on at a constant rate. At the same time, during the third year of life, children engage in role play. The interesting aspect that we noticed is that role play combines with FoVP, although this happens only later in development, that is when role play is a well-established form of play. It is likely

Table 9.2 Rate of subcategories of FoVP



	1st quarter	2nd quarter	3rd quarter	4th quarter
Imp-FoVP	6,4	6,45	5	5,68
Rit-FoVP	1,76	0	2,88	2,53
Pret&FoVP	4,4	2,3	2,5	3,7
Role&FoVP	0	0	0,38	1,89

Quarterly rate (Σ play events/filmed hours) of subcategories of FoVP during the third year

that this developmental sequence that goes from a simple type of play to a combined type of play (i.e., role play by itself to role play combined with FoVP) also appears in relation to pretend play. But since pretend play is present from the beginning of the period we studied, we could not observe whether it first appears in a simple or combined form. Anyhow, the existence of combined forms of play, present from the beginning to the end of the third year, and our *simple-combined sequence hypothesis*, led us to think of Forms of Vitality Play as a good runway for the development of symbolic play.

9.3.3 The Emergence of FoVP: Antecedents During the Second Year of Life

Our data on Forms of Vitality Play is limited to the third year of life. Since it is present from the beginning of the period studied, we are unable to determine when Forms of Vitality Play emerges in ontogenetic development. Our only hint comes from a preliminary study where Bordoni and Español (2018) analyze a free interaction session in a gesture study, between an adult and a 13-month-old baby. Although

no situations qualifying as early social play or as play with the forms of vitality were identified, we recognized "intermediate forms of play" characterized as pleasurable and non-figurative interactive situations, where interpersonal involvement does not result in the fulfillment of practical purposes, and where adult or infant participation modes, or both, are not typical of early social play or forms of vitality play. We found an intermediate form in play interactions where the infant participates with general reciprocity patterns (such as smiling, gazing, and vocalizing) and even with some small actions in response to the play motif created by the adult, but where, however, the child is not involved in the co-construction of the sound-kinetic motif or in its various repetitions as it does happen in FoVP. Another type of intermediate form is created through mutual imitation, where by taking turns in a simple behavioral pattern, the adult proposes elaborations of the imitated behavior not fully matched by the infant, resulting in a less fluid turn-taking organization. There is also an intermediate form where the adult responds to interesting infant initiatives (for example, repetitive and pleasurable movements or vocalizations) with different reciprocity patterns−smiles, glances, body contact, comments, etc. – but without participating in the infant's aesthetic behavior elaboration. These intermediate forms of play – found at the beginning of the second year of life – show that vitality play does not emerge abruptly; instead, there is a gradual transition between early social play and vitality forms of play during the second year of life. On the other hand, in recently initiated studies (García Cernaz et al., under review), forms of vitality play events have been detected later in the second year of life. Further longitudinal studies are needed to observe when forms of vitality play emerges, and whether its early manifestations arise simply or in combination with pretend play, or with functional play which is the preceding pre-symbolic play.

9.4 Final Remarks

The aesthetic perspective we assumed in this book led us to highlight – in Chap. 2 – the non-conscious, non-deliberate improvisational skills arising when an adult comes into contact with a baby. These skills are sustained by a simple strategy at the service of reciprocity: the repetition-variation form. In this final chapter, this aesthetic perspective led us to illuminate the child´s ability to improvise playfully in interpersonal contexts using the same strategy. In FoVP the infant reveals that he/she has been incorporating a composition and improvisation strategy, of which he/she has been an observer or a receiver for at least 2 years.

Expertise in perceiving and generating variations in vitality forms is one of the nonverbal cues of interpersonal encounters. According to Stern (2010), without vitality forms, the exquisite fine-tuning of interpersonal interactions would not occur. FoVP can be seen as a playful activity that favors, firstly, the experiences and recognition of different forms of vitality; and secondly, the elaboration of symbolic events within a risk-free context where a multiple interactive pairing of sound and motor gestures is exerted. During FoVP, the child can experience his/her own

agency or authorship in the vicissitudes of interpersonal encounters, actively entering experiences of intersubjectivity and mutuality that characterize the human world. Together with the adult, she or he co-creates moments of intimacy that may involve attunements or contrasts between forms of vitality. At times adult and baby may agree, for example, on the intensity of sounds and vocalizations, and at other times they may playfully contest intensity (soft versus loud or vice versa). Reciprocity does not necessarily imply an agreement, but a dynamic dialogical construction. In the intersubjective context of FoVP, babies safely rehearse the aesthetic elaboration of vitality forms. In their vocalizations, they can model a fortissimo while not disturbing the mother, father, or accompanying adult, because they shout but do not really shout, they are playing with behavior intensity. A similar yet different thing occurs in the puppy play described at the beginning of this chapter. The "epiphany by denial" that we find in the play of puppies (the animal strips the hurtful bite of its effects by suspending it and turning it into a playful and soft caress involving its teeth) becomes much more complex and diverse in FoVP of human babies. In the epiphany by denial carried by non-human mammals, the message *this is play* is conveyed by a decrease in intensity or suspension of the action. In the human FoVP, instead, the playful aspect is achieved through all kinds of variations of forms of vitality including decrease as well as increase and many others. Play complexity protects and frames the increased intensity and tension. The adult-child complicity guarantees modulation and calmness.

FoVP appears in different ways. Each of them challenges us in a specific fashion. Improvised FoVP, like the Infant-Directed Performances described in Chap. 2, brings to our attention all that is ephemeral, and non-thematic (in narrative terms). Improvised FoVP are fleeting, ephemeral, and beautiful interactive moments. In general, they reach a substantial form and disappear once executed. They may linger from session to session, accompanied by the joy of recognition, but eventually they disappear. On the other hand, ritualized FoVP demonstrates that pre-established exchange forms and cultural elements of the environment, supported by the ability to improvise, integrate into play that focus on the temporal, and spatial sound and movement elaboration, despite presenting some figurative traits. Finally, the combined forms reveal that narrative and symbolic capacities, far from being alien to the varied plot of sounds and movements, are significantly intertwined, as the child uses FoVP to create and elaborate symbolic events and the narrative of the play.

Stern (2010) posits the idea that forms of vitality are a meeting point between early social play and time-based arts. FoVP can be considered an *ontogenetic bridge* between early social play and temporal arts. Like early social play, FoVP requires the support of social interaction, but it is different from that early form of play insofar as it involves the infant's active and specific participation. As temporal arts do, FoVP presupposes the ability to modulate vitality forms by means of expressive resources, like timing or the repetition-variation form; however, FoVP is a type of play and not an artistic expression linked to cultural, biographical, zeitgeist capturing, or some sort of transcendence, like art forms. FoVP may not be the only bridge between early social play and temporal arts, and children most likely rehearse the shaping of the forms of vitality in other contexts as well.

Scribbling is a context as privileged as play. During their second year of life, infants begin to trace lines on surfaces. Although the first scribbles may be made with no visual control, controlled scribbles soon appear, where the awareness of being an agent of movement and author of a product brings the infant obvious pleasure. The first scribbles are not intended to represent formal aspects of reality. There is no reproduction of the formal, static, and geometric qualities of objects. The figurative, *realistic* stroke appears later in development. But from the first months of life, the infant experiences the meaning of variations in the intensity, form, and temporality of the sounds and movements he/she expresses and perceives from others. He/she has repeatedly experienced soft and surprising, strong and continuous forms, or other combination. He/she has experienced direct and sinuous movements, surprising or constant. In a recently initiated observational study that went on between the 24th and 36th month of a girl called Luna, Español observed that in scribbling, the size, speed, and strength of the stroke express different dynamic feelings and sensations. Thus, when Luna is 1 year and 5 months old, she quickly delineates an extended circular line – scribble 17 in Kellogg´s classification (Kellogg, 1969 and Kellogg archive). Once she finishes, she says "Heart". She brings her hand to her chest and taps it several times. She is so happy with what she did that she immediately makes several more scribbles, with fast, strong strokes, and says "this is also a heart", "this one too". She seems to be tracing the dynamic experience of her racing heart. At age 2, she draws several long lines, mostly rounded, but also straight in some parts. The stroke is sometimes fast, sometimes slow, but mostly steady. While drawing her lines, the girl says "a song". The lines seem to express the child's immersive experience with the songs or a particular song. At age 2 and a half, Luna outlines a soft swirling line of Kellogg's scribble 11 type (2021), it is barely noticeable because the stroke is very, very soft, and small; the girl delineates it slowly, bringing her face close to the paper; at the same time she says "water". At 2 years and 10 months, Luna is walking down the street while a strong wind blows, and when she gets back home, while tracing strong, fast, repeated lines overlapped with vibrant red on a paper, she says "wind". The child seems to express the dynamic grasp of the strong wind she has been perceiving in her body and surroundings.

Most likely, children rehearse the molding of vitality forms in other contexts as well. This is an open door for research. Ongoing research on vitality forms in the arts, discussed in point 2, may lead to the recognition of other forms of elaboration and expression of vitality forms in childhood. For example, Stern (2010) underlined the extraordinary new dynamic forms of vitality in the lighting and scenic arrangements created by Robert Wilson. Also Bussières (2018) proposed composing with light, as changes in light intensity trigger emotional shifts and establish atmospheres in a primitive way. Stern (1990) detailed the infant's sensitivity to changes in light intensity. We may have overlooked children´s activities related to light, or may have failed to provide an environment conducive to their emergence. In any case, it is reasonable to imagine that activities where the expression of forms of vitality is given by the dynamics of light could already exist in infant activities (unnoticed by research) or could be easily elicited. Likewise, research on art forms and on play – outlined in points 2 and 3 respectively – suggest that new forms of expression in

child development, both in simple and combined forms, could be detected by future research.

Forms of vitality operate across modalities and elicit similar feeling states, regardless of what modality they arise from, and represent a privileged way for the arts to collaborate with each other. As indicated by Stern (2010), when juxtaposing different art forms, the magic lies in matching the similar with the not-so-similar. As in the arts, pairing similar with not-so-similar shapes of sound and movement in the play scenarios described in point 3 results in the magic of simple FoVP. Combined FoVP where children articulate the modeling of vitality forms with fictional narrative are reminiscent of Thurston and Slee's (2017) proposal, which seek to activate vitality dynamics as compositional and performance principles while exploring the gap relative to language-based arts. We also would like to mention that when we began the study on scribbling, we noticed that in the example of play 3, the infant-adult elaboration of sounds and movements in the construction of improvised forms of vitality play incorporated scribbles (the lines that girl and adult draw alternately on the paper) into the play. Consequently, still undetected forms of vitality expression might possibly combine different activities.

As mentioned earlier, the recognition of human's broad experience with forms of vitality (which relocates basic emotions into a limited space within our psychological life) is perhaps one of the most salient proposals for acknowledging forms of expression in childhood and in the arts. Through this work, we have tried to advance our understanding of the forms of expression in infancy.

References

Bainbridge Cohen, B. (2012). *Sensing, feeling, and action: The experiential anatomy of body-mind centering*. Contact Editions.

Balducci, A. (2020). Forms of vitality and microanalysis in music therapy within adult autism: A clinical report. *Approaches: An Interdisciplinary Journal of Music Therapy Advance Online Publication, 13*, 220.

Barker, J. (2019). Cinema and child's play. In J. Hanich & D. Fairfax (Eds.), *The structures of the film experience by Jean-Pierre Meunier. Historical assessments and phenomenological expansions* (pp. 304–320). Amsterdam University Press.

Bates, E., Benigni, I., Bretherton, L., Camaioni, L., & Volterra, V. (1979). Cognition and communication from nine to thirteen months: Correlational findings. In E. Bates (Ed.), *The emergence of symbols: Cognition and communication in infancy* (pp. 69–140). Academic Press.

Bateson, G. (1972). *Steps to an ecology of mind*. Chandler.

Bell, R. Q. (1953). Convergence: An accelerated longitudinal approach. *Child Development, 24*(2), 145–152. https://doi.org/10.2307/1126345

Bergen, D. (2013). Does pretend play matter? Searching for evidence: Comment on Lillard et al. (2013). *Psychological Bulletin, 139*(1), 45–48. https://doi.org/10.1037/a0030246

Bergen, D. (2018). Infant sensorimotor play: Development of sociocultural competence and enactive cognition. In P. Smith & J. Roopnarine (Eds.), *The Cambridge handbook of play: Developmental and disciplinary perspectives. Cambridge handbooks in psychology* (pp. 125–141). Cambridge University Press. https://doi.org/10.1017/9781108131384.008

Bjørgen, K., & Svendsen, B. (2015). Kindergarten practitioners' experience of promoting children's involvement in and enjoyment of physically active play: Does the contagion of physical energy affect physically active play? *Contemporary Issues in Early Childhood, 16*(3), 257–271. https://doi.org/10.1177/1463949115600025

Bordoni, M., & Español, S. (2011). Moviéndonos juntos: el movimiento en el juego musical imitativo [Moving together: movement in imitative musical play]. In Herrera, R. & Burcet, Ma. I. (Eds.), Proceedings of X Encuentro de Ciencias Cognitivas de la Música: Musicalidad Humana. Debates actuales en evolución, desarrollo, cognición e implicancias socio-culturales, Buenos Aires: SACCoM.

Bordoni, M., & Español, S. (2018). Antecedentes del juego con las formas de la vitalidad al inicio del segundo año de vida [Antecedents of forms of vitality play at the beginning of the second year of life]. In N. Alessandroni, & M. Burcet, La experiencia musical. *Investigación, interpretación y prácticas educativas. Actas del 13° Encuentro de Ciencias Cognitivas de la Música* (pp. 76−89). Buenos Aires: Sociedad Argentina para las Ciencias Cognitivas de la Música (SACCoM).

Bordoni, M., & Martínez, I. C. (2011). Imitación mutua y juego musical en la infancia [Mutual imitation and play in infancy]. *Psicología del desarrollo, 1*, 69–78.

Boston Change Process Study Group. (2010). *Change in psychotherapy: A unifying paradigm*. W. W. Norton & Company.

Bruner, J. S. (1964). The course of cognitive growth. *American Psychologist, 19*(1), 1. https://doi.org/10.1037/h0044160

Bruner, J. S. (1972). Nature and uses of immaturity. *American Psychologist, 27*(8), 687–708. https://doi.org/10.1037/h0033144

Bruner, J. (1990). *Actos de significado* [Acts of meaning]. Alianza.

Bruner, J. S., Sherwood, V., Jolly, A., & Sylva, K. (1976). Early rule structure: The case of peekaboo. In M. Gauvain & M. Cole (Eds.), *Readings on the development of children* (pp. 71–75). Worth Publishers.

Bruschweiler-Stern, N. (2017). The music of Dan's life. *Psychoanalytic Inquiry, 37*(4), 220–227. https://doi.org/10.1080/07351690.2017.1299492

Buchsbaum, D., Bridgers, S., Weisberg, D. S., & Gopnik, A. (2012). The power of possibility: Causal learning counterfactual reasoning, and pretend play. *Philosophical Transactions of the Royal Society, 36*, 2202–2212. https://doi.org/10.1098/rstb.2012.0122

Burghardt, G. M. (2011). Defining and recognizing play. In A. D. Pellegrini (Ed.), *Oxford handbook of the development play* (pp. 83–100). Oxford University Press.

Bussières, N. (2018). *Light, vitality, and dynamism. An introduction of time and movement into theatre lighting design* [Thesis Presented for the Partial Fulfillment of the Requirements For the Degree of Master of Arts (Individualized Program), Concordia University, Montreal].

Casartelli, L., Cesareo, A., Biffi, E., Campione, G. C., Villa, L., Molteni, M., & Sinigaglia, C. (2020a). Vitality form expression in autism. *Scientific Reports, 10*(1), 1–13. https://doi.org/10.1038/s41598-020-73364-x

Casartelli, L., Federici, A., Fumagalli, L., Cesareo, A., Nicoli, M., Ronconi, L., … Sinigaglia, C. (2020b). Neurotypical individuals fail to understand action vitality form in children with autism spectrum disorder. *Proceedings of the National Academy of Sciences, 117*(44), 27712–27718. https://doi.org/10.1073/pnas.2011311117

Di Cesare, G., Di Dio, C., Rochat, M. J., Sinigaglia, C., Bruschweiler-Stern, N., Stern, D. N., & Rizzolatti, G. (2013). The neural correlates of 'vitality form' recognition: An fMRI study. This work is dedicated to Daniel Stern, whose immeasurable contribution to science has inspired our research. *Social Cognitive & Affective Neuroscience, 9*(7), 1918–1924. https://doi.org/10.1093/scan/nst068

Di Cesare, G., Di Dio, C., Rochat, M. J., Sinigaglia, C., Bruschweiler-Stern, N., Stern, D. N., & Rizzolatti, G. (2014). The neural correlates of 'vitality form' recognition: An fMRI study: This work is dedicated to Daniel Stern, whose immeasurable contribution to science has inspired our

research. *Social Cognitive and Affective Neuroscience, 9*(7), 951–960. https://doi.org/10.1093/scan/nst068

Di Cesare, G., Gerbella, M., & Rizzolatti, G. (2020). The neural bases of vitality forms. *National Science Review, 7*(1), 202–213. https://doi.org/10.1093/nsr/nwz187

Dissanayake, E. (2001). *Homo aestheticus: Where art comes from and why*. University of Washington Press.

Duncan, S., & Duncan, T. (2012). Accelerated longitudinal designs. In B. Laursen, T. Little, & N. Card (Eds.), *Handbook of developmental research methods* (pp. 31–45). Guilford Press.

Elkonin, D. B. (1985). *Psicología del Juego* [Play psychology]. Visor.

Español, S. (2004). *Cómo hacer cosas sin palabras. Gesto y ficción en la infancia temprana* [How to do thing without words: Gesture and fiction in early infancy]. A. Machado Libros.

Español, S. (2014). La forma repetición-variación. Una estrategia para la reciprocidad [Variation-repetition form: A strategy for reciprocity]. In S. Español (Ed.), *Psicología de la música y del desarrollo. Una exploración interdisciplinaria sobre la musicalidad humana* (pp. 157–192). Paidós.

Español, S., & Pérez, D. (2015). Los juegos, los contrastes y las artes [Kinds of play, contrasts, and arts]. *Boletín de estética, 33*, 5–40. http://www.boletindeestetica.com.ar/index.php/boletin/issue/view/51/NroCompleto33

Español, S., Martínez, M., Bordoni, M., Camarasa, R., & Carretero, S. (2014). Forms of vitality play. *Integrative Psychological and Behavioral Science*, 1–27. https://doi.org/10.1007/s12124-014-9271-5

Español, S., Bordoni, M., Martínez, M., Camarasa, R., & Carretero, S. (2015). Forms of vitality play and symbolic play during the third year of life. *Infant Behavior and Development, 40*, 242–251. https://doi.org/10.1016/j.infbeh.2015.05.008

Fagen, R. M. (2011). Play and development. In A. D. Pellegrini (Ed.), *Oxford handbook of the development play* (pp. 83–100). Oxford University Press.

Fantasia, V., Fasulo, A., Costall, A., & López, B. (2014). Changing the game: Exploring infants' participation in early play routines. *Frontiers in Psychology, 5*, 522. https://doi.org/10.3389/fpsyg.2014.00522

Feldenkrais, M. (2011). *Embodied wisdom: The collected papers of Moshe Feldenkrais*. North Atlantic Books.

García Cernaz, S., Martínez, I. C., & Español, S. (under review). Musicalidad y juego en la infancia: el andamiaje que el juego con las formas de la vitalidad provee al juego de ficción [Musicality and play in childhood: The scaffolding that forms of vitality play provides to pretend play]. *Interdisciplinaria*.

Garvey, C. (1977). *El juego infantil* [Play]. Morata.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. Appleton-Century-Croft.

Glaser, B. G., & Strauss, A. L. (1967). *The discovery of grounded theory*. Aldine Publishing Company.

Gómez, J.C. (2004/2007). *El desarrollo de la mente en los simios, los monos y los niños* [The development of the mind in apes, monkeys, and children]. Morata.

Göncü, A. (1993). Development of intersubjetivite in social pretend play. *Human Development, 36*, 185–198. https://doi.org/10.1159/000278206

Goncü, A., & Gaskins, S. (2010). Comparing and extending Piaget's and Vygotsky's understandings of play: Symbolic play as individual, sociocultural, and educational interpretation. In P. Nathan & A. D. Pellegrini (Eds.), *The Oxford handbook of the development of play* (pp. 48–57). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195393002.013.0005

Guzdial, M. J. (2015). *Narrative development in collaborative pretend play*. Gerogia Institute of Technology.

Hobson, R. P. (1993). *Autism and the development of mind*. Erlbaum.

Hobson, R. P., & Lee, A. (1999). Imitation and identification in autism. *The Journal of Child Psychology and Psychiatry and Allied Disciplines, 40*(4), 649–659. https://doi.org/10.1111/1469-7610.00481

Hoffman, J., & Russ, S. (2012). Pretend play, creativity, and emotion regulation in children. *Psychology of Aesthetics, Creativity, and the Arts, 6*(2), 175–184. https://psycnet.apa.org/doi/10.1037/a0026299

Iacone, S. (2019). Paraphilias and violence in the couple. An integrative approach. *Mediterranean Journal of Clinical Psychology, 7*(2).

Johnson, M. (2007). *The meaning of the body*. Psychology Press, Taylor & Francis.

Karniol, R., (2016). A language-based three-stage social-interactional model of social pretend play: Acquiring pretend as an epistemic operator pretending that and pretending with (the P–PT–PW model). Developmental Review, 411–437. https://doi.org/10.1016/j.dr.2016.06.001

Kavanaugh, R. (2002). Caregiver-child social pretend play: What transpires? In R. Mitchell (Ed.), *Pretending and imagination in animals and children* (pp. 91–101). Cambridge University Press.

Kavanaugh, R. D. (2011). Origins and consequences of social pretend play. In P. Nathan & A. D. Pellegrini (Eds.), *The Oxford handbook of the development of play* (pp. 296–307). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195393002.013.0022

Kellogg, R. (1969). *Analyzing children´s art*. National Press Book.

Kellogg, R. (2021). *Archive*. https://www.early-pictures.ch/kellogg/archive/en/query?cat=K-AA. Accessed 21 October 2021.

Kim, J. H. (2013). Shaping and co-shaping forms of vitality in music: Beyond cognitivist and emotivist approaches to musical expressiveness. *Empirical Musicology Review, 8*(3-4), 162–173. https://doi.org/10.18061/emr.v8i3-4.3937

Kim, J. H. (2020). From the body image to the body schema, from the proximal to the distal: Embodied musical activity toward learning instrumental musical skills. *Frontiers in Psychology, 11*(101), 1–8. https://doi.org/10.3389/fpsyg.2020.00101

Laban, R. (1971). The Mastery of movement. Boston: Plays Inc.

Lakoff, G., & Johnson, M. (1980). The metaphorical structure of the human conceptual system. *Cognitive Science, 4*(2), 195–208. https://doi.org/10.2307/2025464

Lakoff, G., & Johnson, M. (1999). *Philosophy in the flesh: The embodied mind and its challenge to western thought* (Vol. 640). Basic books.

Langer, S. K. (1954). *Philosophy in a new key: A study in the symbolism of reason, rite, and art*. The New American Library.

Langer, S. K. (1967). *Mind: An essay on human feelings* (Vol. 1). Johns Hopkins University Press.

Lauffenburger, S. K. (2020). 'Something more': The unique features of dance movement therapy/psychotherapy. *American Journal of Dance Therapy, 42*, 16–32. https://doi.org/10.1007/s10465-020-09321-y

Leslie, A. M. (1987). Pretense and representation: The origins of "Theory of Mind". *Psychological Review, 94*(4), 412–426. https://doi.org/10.1037/0033-295X.94.4.412

Leslie, A. M. (2005). Developmental parallels in understanding minds and bodies. *Trends in Cognitive Sciences, 9*(10), 459–462. https://doi.org/10.1016/j.tics.2005.08.002

Lillard, A. S., & Kavanaugh, R. D. (2014). The contribution of symbolic skills to the development of an explicit theory of mind. *Child Development, 85*, 1535–1551. https://doi.org/10.1111/cdev.12227

Lillard, A. S., Lerner, M. D., Hopkins, E. J., Dore, R. A., Smith, E. D., & Palmquist, C. (2013). The impact of pretend play on children's development: A review of the evidence. *Psychological Bulletin, 139*(1), 1–34. https://doi.org/10.1037/a0029321

Malloch, S., & Trevarthen, C. (2008). Musicality: Communicating the vitality and interests of life. In S. Malloch & C. Trevarthen (Eds.), *Communicative musicality* (pp. 1–11). Oxford University Press.

Mandler, J. M. (2005). How to build a baby III: Image-schemas and the transition to verbal thought. In B. Hampe (Ed.), *From perception to meaning: Image schemas in cognitive linguistics* (pp. 137–163). Mouton de Gruyter.

McCune, L. (2010). Developing symbolic abilities. In B. Wagoner (Ed.), *Symbolic transformation: The mind in movement trough culture and society* (pp. 173–192). Routledge Press.

McCune, L., & Agayof, J. (2002). Pretending as representation: a developmental and comparative view. En R. W. Mitchell, Pretending and Imagination in Animals and Children (pp. 43–55). Cambridge: Cambridge university Press.

Meins, E., Fernyhough, C., Arnott, B., Leek, S. R., & Rosnay, M. (2013). Mind-mindedness and theory of mind: Mediating roles of language and perspectival symbolic play. *Child Development, 84*(5), 1777–1790. https://doi.org/10.1111/cdev.12061

Moore, C. L. (2009). *The harmonic structure of movement, music, and dance according to Rudolf Laban: An examination of his unpublished writings and drawings*. Edwin Mellen Press.

Neisser, U. (1993). *The self perceived*. Cambridge University Press.

Nicolopoulou, A., & Ilgaz, H. (2013). What do we know about pretend play and narrative Development? A response to Lillard, Lernes, Hopkins, Dore, Smith, and Palmquist on "the impact of Preten play on children's development: A review of the evidence". *American Journal of Play, 6*(1), 55–81.

Nudler, A., Español, S., Jacquier, P., & Porcel de Peralta, A. (2018). A multimodal analysis of vitality forms in the play Krapp's Last Tape. In R. Parncutt, & S. Sattmann (Eds.), *Proceedings of ICMPC15/ESCOM10* (pp. 341−343). Graz, Austria: Centre for Systematic Musicology, University of Graz. https://static.uni-graz.at/fileadmin/veranstaltungen/music-psychology-conference2018/documents/ICMPC15_ESCOM10%20Proceedings.pdf

Orr, E., & Geva, R. (2015). Symbolic play and language development. *Infant Behavior and Development, 38*, 147–161. https://doi.org/10.1016/j.infbeh.2015.01.002

Papoušek, M. (1996). Intuitive parenting: A hidden source of musical stimulation in infancy. In I. Deliège & J. Sloboda (Eds.), *Musical beginnings. Origins and development of musical competence* (pp. 88–112). Oxford University Press.

Paxton, S. [Culturgest]. (2020, April 9). Steve Paxton conference at Culturgest, Lisbon 2019 [Video]. YouTube. https://youtu.be/LFn8JuCkVrM

Pederson, L., & Moe, V. F. (2020). The moment of play and movement: A qualitative study of children's playful shared movements. *Journal for Research in Arts and Sports Education, 4*(2), 68–83. https://doi.org/10.23865/jased.v4.2218

Pellegrini, A. D. (2009). *The role of play in human development*. Oxford University Press.

Pellegrini, A. D. (2013). Play. In P. D. Zelazo (Ed.), *The Oxford handbook of developmental psychology, Vol. 2. Self and other* (pp. 276–299). Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199958474.001.0001

Pellegrini, A. D., & Smith, P. K. (1998). Physical activity play: The nature and function of a neglected aspect of play. *Child Development, 69*(3), 577–598. https://doi.org/10.1111/j.1467-8624.1998.tb06226.x

Piaget, J. (1946/2012). *La formación del símbolo en el niño* [The formation of the symbol in the child]. Fondo de Cultura Económica.

Postacchini, P. L., Ricciotti, A., & Borghesi, M. (2014). *Musicoterapia [Music therapy]* (2nd ed.). Carocci editore.

Rapko, J. (2020, February 17). Vitality in the arts. Anne Teresa de Keersmaeker's Phase in light of Daniel Sterns analysis of forms of vitality. *Return to darkness. A dirty, poorly-organized place for philosophical remarks on the arts*. https://johnrapko.com/blog/2020/2/17/vitality-in-the-arts-anne-teresa-de-keersmaekers-phase-in-light-of-daniel-sterns-analysis-of-forms-of-vitality. Accessed 23 May 2021.

Reddy, V. (2008). *How infants know minds*. Harvard University Press.

Riba, C. (1990). *La comunicación animal. Un enfoque zoosemiótico* [Animal communication. A zoosemiotic approach]. Anthropos.

Rivière, A. (1998). Tratamiento y definición del espectro autista I: relaciones sociales y comunicación [Treatment and definition of the autism spectrum I: Social relations and communication]. In A. Rivière & J. Martos (Eds.), El tratamiento del autismo. Nuevas Perspectivas [The treatment of autism. New perspectives] (pp. 61−106). IMSERSO.

Rivière, A. (1984/2003a). Acción e interacción en el origen del símbolo [Action and interaction at the origin of the symbol]. In A. Rivière (Ed.), *Obras Escogidas*. *Lenguaje, simbolización y alteraciones del desarrollo*, *Vol. II* (pp. 77−108). Editorial Médica Panamericana.

Rivière, Á. (2003b). El juego simbólico en niños ciegos [Symbolic play in blind children]. In E. M. Belinchón, A. Rosas, M. Sotillo, & I. Marichalar (Eds.), *Ángel Rivière. Obras escogidas. Volumen II. Lenguaje, simbolización y alteraciones del desarollo* (pp. 173–192). Médica Panamericana.

Rivière, Á., & Español, S. (2003). La suspensión como mecanismo de creación semiótica [Suspension as a semiotic creation mechanism]. *Estudios de Psicología, 24*(3), 261–275. https://doi.org/10.1174/021093903770411184

Rochat, P. (2000/2004). *El mundo del bebé* [The infant´s world]. Morata.

Rochat, M. J., Veroni, V., Bruschweiler-Stern, N., Pieraccini, C., Bonnet-Brilhault, F., Barthélémy, C., Malvy, J., Sinigaglia, C., Stern, D. N., & Rizzolatti, G. (2013). Impaired vitality form recognition in autism. *Neuropsychologia, 51*(10), 1918–1924. https://doi.org/10.1016/j.neuropsychologia.2013.06.002

Rosch, E. (1978). Principles of categorization. In E. Rosch & B. Lloyd (Eds.), *Cognition and categorization* (pp. 27–48). Lawrence Erlbaum.

Sansanwal, S. (2014). Pretend play enhances creativity and imagination. *Journal of Arts and Humanities, 3*(1), 70–83. https://doi.org/10.18533/journal.v3i1.340

Sheets-Johnstone, M. (2012). From movement to dance. *Phenomenology and the Cognitive Sciences, 11*(1), 39–57. https://doi.org/10.1007/s11097-011-9200-8

Shifres, F. (2008). Música, transmodalidad e intersubjetividad [Music, transmodality and intersubjectivity]. *Estudios de Psicología, 29*(1), 7–30. https://doi.org/10.1174/021093908783781347

Shifres, F., Pereira Ghiena, A., Herrera, R., & Bordoni, M. (2012). Estilo de ejecución musical y de danza en el tango: atributos, competencia y experiencia dinámica [Musical and dance performance style in tango: Attributes, competence and dynamic experience]. *Cuadernos de música, artes visuales y artes escénicas, 7*(2), 83–108.

Smith, L. B. (2013). It's all connected: Pathways in visual object recognition and early noun learning. *American Psychologist, 68*(8), 618–629. https://doi.org/10.1037/a0034185

Smith, L. B., & Jones, S. S. (2011). Symbolic play connects to language trough visual object recognition. *Developmental Science, 14*(5), 1142–1149. https://doi.org/10.1111/j.1467-7687.2011.01065.x

Stern, D. N. (1971). A micro-analysis of mother-infant interaction. Behavior regulating social contact between a mother and her 3 1/2 month-old twins. *Journal of the American Academy of Child Psychiatry, 10*(3), 501–517. https://doi.org/10.1016/s0002-7138(09)61752-0

Stern, D. (1973). "On Kinesic analysis": A discussion with Daniel N. Stern. *The Drama Review*: *TDR*, *17*(3) Theatre and the Social Sciences, 114−126. http://www.jstor.org/stable/1144848

Stern, D. N. (1974). Mother and infant at play: The dyadic interaction involving facial, vocal, and gaze behaviors. In M. Lewis & L. A. Rosenblum (Eds.), *The effect of the infant on its caregiver*. Wiley-Interscience.

Stern, D. N. (1985). *The interpersonal world of the infant: A view from psychoanalysis and developmental psychology* (1st ed.). Basic Books.

Stern, D. N. (1990). *Diary of a baby: What your child sees, feels, and experiences*. Basic Books.

Stern, D. N. (2000). Putting time back into our considerations of infant experience: A microdiachronic view. *Infant Mental Health Journal: Official Publication of the World Association for Infant Mental Health, 21*(1-2), 21–28.

Stern, D. N. (2004). *The present moment in psychotherapy and everyday life (Norton series on interpersonal neurobiology)*. W. W. Norton & Company.

Stern, D. N. (2010). *Forms of vitality: Exploring dynamic experience in psychology, the arts, psychotherapy and development*. Oxford University Press.

Stern, D. N., Beebe, B., Jaffe, J., & Bennet, S. L. (1977). The infant's stimulus world during social interaction: A study of caregiver behaviors with particular reference to repetition and timing. In H. R. Schaffer (Ed.), *Studies in mother-infant interaction* (pp. 177–202). Academic.

Stern, D. N., Bruschweiler-Stern, N., Harrison, A. M., Lyons-Ruth, K., Morgan, A. C., Nahum, J. P., Sander, L., & Tronick, E. Z. (1998a). The process of therapeutic change involving implicit knowledge: Some implications of developmental observations for adult psychotherapy. *Infant Mental Health Journal, 19*(3), 300–308.

Stern, D. N., Sander, L. W., Nahum, J. P., Harrison, A. M., Lyons-Ruth, K., Morgan, A. C., Bruschweiler-Stern, N., & Tronick, E. Z. (1998b). Non-interpretive mechanisms in psychoanalytic therapy. The 'something more' than interpretation. The process of change study group. *International Journal of Psychoanalysis, 79*(5), 903–921.

Sung, J., & Hsu, H. C. (2014). Collaborative mother-toddler communication and theory of mind development at age 4. *Journal of Applied Developmental Psychology, 35*, 381–391. https://doi.org/10.1016/j.appdev.2014.06.001

Thibodeau, R. L., Gilpin, A. T., Brown, M. M., & Meyer, B. A. (2016). The effects of fantastical pretend-play on the development of executive functions: An intervention study. *Journal of Experimental Child Psychology, 145*, 120–138. https://doi.org/10.1016/j.jecp.2016.01.001

Thurston, S., & Slee, S. M. (2017). Vital signs: Poetry, movement and the writing body. *Choreographic Practices, 8*(1). https://doi.org/10.1386/chor.8.1.9_1

Trevarthen, C. (2011). The generation of human meaning: How shared experience grows in infancy. In A. Seemann (Ed.), *Joint attention: New developments in psychology, philosophy of mind, and social neuroscience* (pp. 73–113). MIT Press.

Trevarthen, C. (2017). Play with infants: The impulse for human story-telling. In T. Bruce, P. Hakkarainen, & M. Bredikyte (Eds.), *The Routledge international handbook of play in early childhood* (pp. 198–216). Taylor & Francis/Routledge.

Vermes, K. (2011). Intersensory and intersubjective attunement: Philosophical approach to a central element of dance movement psychotherapy. *Body, Movement and Dance in Psychotherapy, 6*(1), 31–42. https://doi.org/10.1080/17432979.2010.533823

Vlietstra, A. G. (1978). Exploration and play in preschool children and young adults. *Child Development, 49*(1), 235–238. https://doi.org/10.2307/1128616

Vygotski, L. S. (1933/1979). El juego y su papel en el desarrollo psicológico del niño [Play and its rol in infant psychological development]. In L. S. Vygotski (Ed.), *El desarrollo de los procesos psicológicos superiores* (pp. 141−158). Crítica.

Wallace, C. E., & Russ, S. (2015). Pretend play, divergent thinking, and math achievement in girls: A longitudinal study. *Psychology of Aesthetics, Creativity, and the Arts, 9*(3), 296–305. https://doi.org/10.1037/a0039006

Weeks, N. (2013). The dynamics of embodied gestures in the theatre of Robert Wilson. *Interdisciplinary network conference on performance* [Conference session]. Mansfield College, Oxford, UK.

Wengrower, H. (2021). Dance comes to the front stage in dance movement therapy. In H. Wengrower & S. Chaiklin (Eds.), *Dance and creativity within dance movement therapy. International perspectives* (pp. 9–24). Routledge.

Wieder, S. (2017). The power of symbolic play in emotional development through the DIR Lens. *Top Lang Disorders, 37*(3), 259–281. https://doi.org/10.1097/TLD.0000000000000126

Withebread, D., & O'Sullivan, L. (2012). Preschool children's social pretend play: Supporting the development of metacommunication, metacognition and self-regulation. *International Journal of Play, 1*(2), 197–213. https://doi.org/10.1080/21594937.2012.693384

Wojciehowski, H. (2014). Statues that move: Vitality effects in the winter's Tale. *Literature & Theology, 28*(*3*), 1–17. https://doi.org/10.1093/litthe/fru022

Žvelc, G. (2012). Mindful processing in psychotherapy: Facilitating natural healing process within attuned therapeutic relationship. *International Journal of Integrative Psychotherapy, 3*(1), 42–58.

Žvelc, G., & Žvelc, M. (2008, July 1−4). *The power of present moment. Mindful processing in psychotherapy and counseling [Workshop]*. In 4th European conference on positive psychology, Rijeka, Opatija, Croatia.

Silvia Español holds a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). She is a currently researcher at the Latin America Institute of Social Research (IICSAL), Faculty of Latin America Social Sciences/National Scientific and Technical Research Council (FLACSO/CONICET, Argentina). She coordinates the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc. in Cognitive Psychology, and Director of FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is a certified Somatic Educator in the Feldenkrais Method. Her work deals with cognitive developmental psychology, psychology of music, and the study of human movement. She has written numerous papers in her area of expertise: https://www.aacademica.org/silvia.espanol

Mauricio Martínez has a PhD. in Psychology from Universidad Autónoma de Madrid (Spain). He participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning at the Faculty of Latin America Social Sciences (FLACSO). He is Professor of Developmental Psychology at Universidad Abierta Interamericana (UAI) and Faculty member of the MsC. in Cognitive Psychology (FLACSO). He also teaches in the post-graduate virtual course The baby's life: psychological development, scaffolding and nurturing environment" at FLACSO. His research focuses on early perceptual and intersubjective development, childhood play and early intervention in autism spectrum disorders. He is author of Intervención psicoeducativa para niños con trastorno del espectro autista [Psychoeducational intervention for children with autism spectrum disorders] (2015). His publications are available at: https://www.researchgate.net/profile/Mauricio-Martinez-2

Mariana Bordoni has a PhD. in Psychology from Universidad Nacional de Córdoba (Argentina). She participates in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). She is a Faculty member of the MSc. in Cognitive Psychology at Universidad de Buenos Aires (UBA) and FLACSO. She also teaches in FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. She is the creator and coordinator of the Baby Social Club. She has collaborated in several research projects funded by UBACyT and ANCPyT. Her publications are available at: https://www.aacademica.org/mariana.bordoni

Rosario Camarasa is a licensed psychologist specializing in child psychoanalysis. She has worked in the area of child mental health in public hospitals, clinics and in her private practice in Buenos Aires. She is currently a researcher at Universidad Abierta Interamericana (UAI, Argentina). She teaches in the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO). Her interests focus on early childhood and infant development from both cognitive and psychoanalytic perspectives. Her publications are available at: https://uai-ar.academia.edu/RosarioCamarasa

Alicia Nudler is a psychologist from Universidad de Buenos Aires, Argentina, and Master of Education from University of Massachusetts, United States. She is a Professor at Universidad de Río Negro in Patagonia, Argentina, where she teaches Psychology to undergraduate drama students. Sghe works at the Laboratory of Research in Education of Teachers, Universidad de Río Negro. She is a doctoral candidate in the Theory and History of Arts Program in Universidad de Buenos Aires. Her research is in the field of theatre studies, cognitive embodied psychology and forms of vitality. Her case of study is Krapp's Last Tape by Samuel Beckett, where she looks at specific performances of the play, both in English and Spanish, from the point of view of embodied cognition and forms of vitality. She coordinates the Research Project *Dynamic forms of vitality in the performing arts: A study from embodiment* at Universidad de Rio Negro. She also carries an amateur practice in contemporary dance. Her publications are available at: https://www.unrn.academia.edu/alicianudler

Santiago García Cernaz is a psychologist from Universidad Nacional del Comahue (UNCo) and a PhD. candidate at the Universidad Nacional de La Plata (UNLP, Argentina). He is fellow of the National Scientific and Tecnological Research Council (CONICET) at the Laboratory for the Study of Musical Experience (LEEM) in the Faculty of Arts at the UNLP. He is a Faculty member of the UNLP and also a member of the Research Group on Socio-Cognitive Processes in Infancy and Early Childhood at Psychology of Knowledge and Learning Department (FLACSO, Argentina). He works as form tutor in FLACSO's virtual post-graduate course *The baby's life: psychological development, scaffolding and nurturing environment*. His line of work focuses on child socio-cognitive development, particularly play and communicative musicality, in the border area between developmental psychology and the psychology of music.

Index

A
Action song, 13, 51, 58–83, 95, 137, 304
Aesthetic, 6, 7, 9, 12, 13, 17, 22, 26–30, 33, 50, 51, 81, 259, 287, 297, 304, 316, 317
Affect attunement, 12, 14, 30, 92–114, 169, 188, 191, 196, 198, 295, 297, 300, 301
Amodal information, 11, 15, 126, 151, 161–199, 306

B
Bimodal combination, 15, 216, 217, 219–224, 227–238, 241, 242, 257, 266, 271, 272

C
Communicative functions, 16, 248–281
Communicative musicality, 6, 13, 14, 24, 27, 28, 59, 64, 65, 68, 72, 79, 80, 137, 213, 295

D
Dance, 6, 7, 9, 12, 24–26, 28–31, 33, 34, 50, 51, 59, 62, 63, 65, 70, 78, 97, 111, 164, 190, 291, 295–298, 301, 304, 306, 310
Direct social perception, 136, 138
Duration, 11, 15, 23, 26, 27, 31, 34, 38, 39, 41, 42, 47, 60, 67, 70, 71, 73–75, 81, 94, 95, 97, 98, 104, 124, 137, 138, 140, 146, 151, 164–169, 171, 174, 177, 183–187, 189, 191, 193–199, 239, 277, 290, 295, 306–308, 311

E
Early social play, 28, 49, 97–99, 102, 103, 106–108, 110–114, 128, 137, 139, 150, 151, 188–192, 194, 196, 286–291, 293, 295, 296, 302, 303, 312, 316, 317
Embodiment, 2–5, 7, 58
Emotional expressions, 5, 96
Experience, 22, 58, 93, 165, 210, 286
Expression, 24, 59, 96, 171, 208, 289
Expressive resources, 16, 30, 43, 46–50, 65, 226, 262, 290, 317

F
Forms of vitality, 6, 7, 10, 13, 16, 28, 33, 34, 47, 140, 286, 287, 290–301, 304, 309, 312, 316–319
Forms of vitality play, 16, 17, 51, 286–319
Functional reorganization, 163, 179, 182, 184–189, 194

G
Gesture, 22, 67, 96, 169, 208, 297

I
Image-schemas, 13, 35, 36, 43–46, 49, 65, 81, 289, 290
Imitation, 14, 30, 49, 64, 79, 83, 92–114, 132, 133, 137, 140, 169, 191, 278, 295, 311, 316
Improvisation, 27, 29, 35, 40, 47, 51, 63, 64, 67, 297, 304, 316

© The Editor(s) (if applicable) and The Author(s), under exclusive license to Springer Nature Switzerland AG 2022

S. Español et al. (eds.), *Moving and Interacting in Infancy and Early Childhood*,
https://doi.org/10.1007/978-3-031-08923-7

Infant-directed improvised performance, 13, 16, 22–52, 58–83, 94, 95, 137, 302–304
Intersensory development, 183
Intersubjectivity, 2, 7–10, 14, 16, 22–24, 33, 61, 65, 73, 79–81, 97–99, 111, 112, 114, 125–134, 150, 151, 217, 250, 286, 290, 292, 317

K
Kinesphere, 32, 33, 38–40, 47

M
Matching activities, 14, 96, 98, 99, 101, 107, 108, 110–112, 114
Maternal responsiveness, 14, 95, 96, 102, 110, 113, 114
Movement, 2, 5–7, 9, 11–17, 24–50, 59, 60, 62–64, 66, 67, 69, 71–82, 93, 94, 96, 98, 100, 104, 105, 108–110, 112–114, 128, 130, 131, 137, 138, 140–143, 145–150, 162, 165–168, 170, 175, 177, 182–185, 193, 196, 208, 211–216, 220, 223, 226, 230, 231, 234, 238–240, 248, 260, 261, 266, 271, 273–278, 280, 288–294, 296–313, 316–319
Movement temporal coordination, 92–114
Multisensory perception, 3, 11, 14, 15, 93, 138, 141–143, 148–151, 161–199
Music, 2, 6, 12, 23, 26–31, 50, 51, 59, 62–65, 78, 98, 111, 124, 164, 291, 293, 295–297, 299, 300, 304, 309
Musicality, 16, 27, 30, 40, 49, 52, 58, 59, 64, 67, 70–72, 74–77, 80, 213, 240, 242, 248–281, 288, 295

P
Perceptual skill, 14, 15, 124, 126, 136, 143–145, 148, 165–168, 173, 178
Precommunicative functions, 281
Primary metaphor, 13, 35, 36, 43–46, 49, 65, 81
Protoconversation, 13, 24, 29, 49, 51, 58–83, 101, 110, 128, 269

R
Rate, 11, 23, 34, 98, 99, 105, 107, 108, 110, 111, 138, 141, 145, 151, 164–168, 171, 174, 177, 184–187, 189–191, 193, 194, 196–199, 306, 312–315
Reciprocity, 3, 9, 10, 14, 25, 26, 28–31, 36, 50–52, 62, 76, 81, 92–114, 171, 172, 189, 296, 316, 317
Repetition-variation, 13, 16, 23, 26, 28, 29, 34, 37–43, 46–48, 51, 59, 60, 62, 63, 66, 67, 76, 81, 140, 149, 169, 177, 188–190, 192–196, 198, 290, 291, 296, 302–305, 308, 310, 313, 316, 317
Rhythm, 11, 14, 15, 23, 27, 28, 34, 61, 63, 67, 71, 73, 75, 77, 78, 80, 93, 94, 96, 97, 100, 104, 124–151, 162, 164–169, 171, 177, 184–189, 191, 193, 194, 196–199, 276, 278, 290, 295, 306

S
Second person, 2, 3, 7–10, 12, 14–16, 25, 80, 92–94, 125, 126, 134, 135, 217, 256, 296, 297, 302, 303
Semantic roles, 208, 228–237, 240
Sensorimotor play, 286, 288, 289, 291
Social cognition, 2, 8, 12, 33, 49, 79, 92, 134, 136, 290
Social perception, 161–199
Symbolic play, 16, 17, 291, 292, 304, 312–316

T
Temporal arts, 6, 12, 13, 26, 28–30, 34, 46, 51, 59, 60, 62–64, 94, 290, 295, 296, 303, 317
Temporal information, 15, 34, 138, 140, 162, 164, 165, 167, 168, 170, 171, 174, 177, 184, 186, 187, 189, 195, 196

V
Vocalization, 11, 12, 14, 16, 22, 24, 25, 28, 29, 39, 42, 43, 48, 59–64, 67, 71–77, 81, 82, 93–96, 99–102, 104, 105, 108–110, 112–114, 127, 128, 130, 137, 140, 143, 146, 147, 197, 208, 215–217, 225–227, 230, 236, 237, 240, 248–281, 290, 294, 302, 307–311, 316, 317