HAL
open science

Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics

Hélène Cochet, Michèle Guidetti

▶ **To cite this version:**

Hélène Cochet, Michèle Guidetti. Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics. Frontiers in Psychology, 2018, 9, 10.3389/fpsyg.2018.01992 . hal-01996824

HAL Id: hal-01996824

https://hal.science/hal-01996824v1

Submitted on 28 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

frontiers in Psychology

REVIEW
published: 19 October 2018
doi: 10.3389/fpsyg.2018.01992

Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics

Hélène Cochet and Michèle Guidetti*

CLLE, Université de Toulouse, CNRS, UT2J, Toulouse, France

Children exchange information through multiple modalities, including verbal communication, gestures and social gaze and they gradually learn to plan their behavior and coordinate successfully with their partners. The development of joint attention and joint action, especially in the context of social play, provides rich opportunities for describing the characteristics of interactions that can lead to shared outcomes. In the present work, we argue that human–robot interactions (HRI) can benefit from these developmental studies, through influencing the human's perception and interpretation of the robot's behavior. We thus endeavor to describe some components that could be implemented in the robot to strengthen the feeling of dealing with a social agent, and therefore improve the success of collaborative tasks. Focusing in particular on motor precision, coordination, and anticipatory planning, we discuss the question of complexity in HRI. In the context of joint activities, we highlight the necessity of (1) considering multiple speech acts involving multimodal communication (both verbal and non-verbal signals), and (2) analyzing separately the forms and functions of communication. Finally, we examine some challenges related to robot competencies, such as the issue of language and symbol grounding, which might be tackled by bringing together expertise of researchers in developmental psychology and robotics.

Keywords: human–robot interaction, human development, joint attention, joint action, coordination, complexity, gestures

OPEN ACCESS

Edited by:
Jill Popp,
The LEGO Foundation, Denmark

Reviewed by:
Nicolas Cuperlier,
Université de Cergy-Pontoise, France
Gautier Durantin,
The University of Queensland, Australia
Eiji Uchibe,
Advanced Telecommunications Research Institute International (ATR), Japan

****Correspondence:***
Hélène Cochet
helene.cochet@univ-tlse2.fr

Specialty section:
This article was submitted to Developmental Psychology, a section of the journal Frontiers in Psychology

Received: *13 October 2017*
Accepted: *28 September 2018*
Published: *19 October 2018*

Citation:
Cochet H and Guidetti M (2018) Contribution of Developmental Psychology to the Study of Social Interactions: Some Factors in Play, Joint Attention and Joint Action and Implications for Robotics. Front. Psychol. 9:1992. doi: 10.3389/fpsyg.2018.01992

INTRODUCTION

Developmental psychologists aim at describing and explaining changes across the life span in a wide range of areas such as social, emotional, and cognitive abilities. Focusing on childhood is a way of grasping numerous changes, especially in terms of communication: infants gradually learn to identify the common ground they have with others and engage in social interactions. The development of such abilities relies on the personal experiences shared between partners in specific contexts (Liebal et al., 2013), among which social play may offer particularly rich opportunities for children to acquire joint action and joint attention skills. Studying the different forms and functions of communication in this context paves the way for identifying the necessary ingredients

for effective joint activities and therefore better understanding the architecture of human–social interactions. Even though the concept of *effectiveness* may cover different theoretical frameworks, the latter objectives have several applications, for example in supporting children with atypical development, especially when they have difficulty communicating both verbally and non-verbally (e.g., children with autism spectrum disorders, ASD), but also in the field of artificial intelligence. The role of robots in society raises indeed a lot of debates and challenges, as they share more and more space and tasks with humans, for instance in service robotics to assist elderly people. The robots' ability to initiate and respond to social interactions is one of the key factors that will shape their integration in our everyday life in the future. Researchers in social robotics have been working on the question of joint action for over two decades now, sometimes in collaboration with developmental psychologists (e.g., Scassellati, 2000), in order to improve robots' motor and communicative skills. Developmental models of human communicative behavior can indeed help define the components to implement in human–robot interactions (HRI), so as to build rich and natural joint activities (Breazeal et al., 2004; Lemaignan et al., 2017).

The objective of this paper is twofold. First, we intend to present the point of view and some research perspectives of developmental psychologists on joint attention and joint action, in particular in the context of social play. To this end, we will also define, starting from studies on non-human primates, what can be regarded as complex (or rich) and natural (or effective) interactions in both human communication and HRI. Second, we aim to show the extent to which the above-mentioned issues may be of interest to roboticists, in helping conceptualize and implement some variables associated with joint attention and joint action in the context of HRI. Collaborative tasks involving robot and human partners, regarded as tantamount to children's social play, will thus be considered through the prism of pragmatic communication, allowing researchers to dissociate the forms and the functions of communication.

HOW DOES COMMUNICATION DEVELOP IN THE CONTEXT OF SOCIAL PLAY?

The definitions of play include a wide range of activities, which makes it difficult to determine where play begins and where it ends, even though it is traditionally associated with positive affective valence (Garvey, 1990). Play, which occurs in several animal species (most notably in mammals), has been argued to allow "practice of real-world skills in a relatively safe environment" (Byrne, 2015). We will focus here on social play in human children, which may also enable them, as highlighted by Bruner (1973), to "learn by doing" as they interact with one or several partners. At the individual level, children can indeed explore and enhance specific skills like motor control and creativity, while developing for example cooperation abilities at the social level. The concepts of artifact-mediated and object-oriented action, originally formulated by Vygotsky (1999), are particularly relevant to describe these situations: the relationship between the child and the surrounding objects is indeed mediated by cultural means, tools, and signs. Studying the development of play can therefore reveal how children come to represent and think about their environment.

Social attention is a crucial capacity for the emergence of these play situations, allowing children to focus on some of the other's characteristics such as the facial expressions, gaze direction, gestures, and vocalizations. When the direction of another's attention has been identified (for example through gaze following or point following), we can shift our own attention to focus at the same time on the same external object or event as our partner. This process of *joint attention* is usually inferred from behavioral cues, including mainly gaze alternation between one's partner and a specific referent (Bourjade, 2017). Joint attention seems therefore necessary for individuals to perform joint action, i.e., to coordinate their actions in space and time to produce a joint outcome, whether it involves here symbolic play (with or without objects), construction toys, board games or any other forms of play.

Joint attention and joint action begin to appear at the end of the first year in human development (Carpenter et al., 1998), gradually allowing children to integrate the notion of common ground and engage in social interactions. The development of gaze understanding, which has been widely studied, plays a key role in this regard. It was for example shown in a study using habituation-of-looking-time procedure that infants start to understand ecologically valid instances of social gaze between two adults interacting, and to have expectations concerning gaze target at 10 months of age (Beier and Spelke, 2012). Besides, responsive joint attention skills (e.g., gaze following and point following) have been reported to emerge before initiative joint attention skills, from 8 months of age (Corkum and Moore, 1998; Beuker et al., 2013).

However, depending on the authors, the definitions of these social-cognitive skills can be more or less demanding, the main difference lying in whether or not individuals have mutual understanding of their shared focus of attention. The ability to "know together" that we are attending to the same thing as our partner has sometimes been referred to as *shared attention* (Emery, 2000; Shteynberg, 2015), which would develop in parallel with *shared intentionality* (Tomasello and Carpenter, 2007). The latter involves the motivation to share goals and intentions with the other, as well as forms of cognitive representation for doing so. This ability has been argued to constitute a hallmark of the human species (Tomasello et al., 2005), even though it is particularly difficult to assess when verbal language is not available as a clue to these representations (in pre-linguistic children or non-human primates). Similarly, joint action may rely solely on the learning of the cues that appear significant (e.g., gestures and eye contact) to coordinate actions in space and time with a partner, or it may also involve, in a more demanding perspective, the common and explicit knowledge of the objectives of the activity and of the way to achieve them (Tomasello and Carpenter, 2007).

Joint attention and joint action, whether they are accompanied or not with shared and explicit intentions, thus allow children

to participate with others in collaborative activities in which each partner benefits from the joint outcome and/or from the interaction in itself. In a series of experiments, the ability to coordinate with a partner in social games was shown to significantly improve between 18 and 24 months of age, whether the games involved complementary or similar roles (Warneken et al., 2006). In the first game of this study, one person had to send a wooden block down one of a tube mounted on a box on a 20 degrees incline, while the other person had to catch it at the other end with a tin can that made a rattling sound. Two tubes were mounted in parallel so that individuals could perform in turn the different roles. In the second game, two persons had to make a wooden block jump on a small trampoline (67 cm diameter ring covered with cloth) by holding the rim on opposite sides. The trampoline collapsed when being held on only one side. Children successfully participated in both games, although the 24 month-olds were more proficient than the 18 month-olds, and they all produced at least one communicative attempt to reengage the adult partner when the latter stopped participating in the activity. Children for example pointed at the object, and/or vocalized while looking at the adult, which was regarded as evidence for a uniquely human form of cooperation, involving shared intentionality (Warneken et al., 2006). A less "mentalistic" interpretation could be proposed (D'Entremont and Seamans, 2007), but these results nevertheless highlight children's motivation for reinstating joint action toward a shared goal. The development of this capacity has received much attention from researchers, as the initiation of joint attention appears to be strongly related to language comprehension and production in the second and third year of life (Colonnesi et al., 2010; Cochet and Byrne, 2016), as well as to theory of mind ability (e.g., Charman et al., 2000; Milward et al., 2017) in both typical and atypical development (e.g., Adamson et al., 2017).

In addition, the observation of children's behavior during collaborative activities may lead to a thorough description of multimodal communication (e.g., gaze, facial expressions, gestures, and verbal language) and of the way its components become coordinated. For example, the production of gestures gradually coordinates with gaze in the course of development. Children start to produce pointing gestures to orient the attention of another person around 12 months of age; an object, a person or an event can become the shared focus of attention but then children do not usually look at their partner while they point (Franco and Butterworth, 1996). A couple of months later, they are able to alternate their gaze between their partner and the object of interest, which represents a key feature of intentional triadic interactions (Cochet and Vauclair, 2010). At 16 months of age, gaze toward the adult can precede the production of pointing (Franco and Butterworth, 1996), suggesting that children may thus take into account the partner's attentional state before initiating communication (Lamaury et al., 2017).

Children also gradually learn to take account of their partner's facial expressions to infer their emotional state and adjust their response accordingly. Infants are sensitive to the characteristics of faces from very early on; newborns look for example significantly longer at happy expressions than at fearful ones, demonstrating some discrimination skills (Farroni et al., 2007). The still-face paradigm, initially designed by Tronick et al. (1978) also suggests that infants have expectations about interactional reciprocity from a few months of age, partly relying on emotional expression. This sensitivity manifests itself in specific behavioral and physiological responses (e.g., reduced positive affect and gazing at the parent, increased negative affect, rise in facial skin temperature) when the mother puts on a neutral and unresponsive face, after a period of spontaneous play with his/her infant (Aureli et al., 2015). The ability to recognize and identify facial expressions of basic emotions further develops in preschool children, before they can understand a few months later the external causes of emotions and then, around 5 years of age, the role of other's desires or beliefs in emotional expression (Pons et al., 2004).

During play interactions, being attentive to the other's facial expressions allows each partner to consider the emotional nature of the signals (e.g., joy, surprise, and frustration) and to possibly modify his/her own behavior to change or maintain this emotional state. The development of facial expression perception thus plays a key role in the emergence of joint actions, in coordination with other communicative modalities. Facial expressions are indeed usually synchronized with vocalizations and/or gestures, and this from infancy.

The vocal and the gestural modalities also become more and more coordinated as children grow older, which represents a key feature of human communication as we use gestures as we speak throughout our life. Communicative gestures are first complemented by vocalizations, whose prosodic patterns may already code for semantic and pragmatic functions (Leroy et al., 2009). In the second year of life, children then produce their first gesture-word combinations, which have an important role in the transition to the two-word stage (e.g., Butcher and Goldin-Meadow, 2000). Pointing and conventional gestures (e.g., waving goodbye, gestural agreement, and refusal: Guidetti, 2002, 2005) remain in the child repertory after the two-word stage, but other forms of gestural-vocal coordination are observed from 3 years of age with the emergence of co-speech gestures. Although we are usually not aware of producing or perceiving them, co-speech gestures can lend rhythm, emphasize speech and sometimes serve deictic or iconic functions. The deictic presentation of pointing gesture can for example be combined with *vocal pointing*, performed through syntactic or prosodic means (Lœvenbruck et al., 2008). Such coordination between the vocal and gestural modalities is omnipresent in adults and play a crucial role in face-to-face communication for both speaker and listener (e.g., McNeill, 2000; Kendon, 2004).

Moreover, the characteristics of gaze, gestures, and vocalizations and their coordination may vary according to the communicative function of the signal. A gesture can indeed serve different purposes, starting with the traditional distinction between imperative and declarative functions (Bates et al., 1975). Imperative gestures are used to request a specific object or action from a partner whereas declarative gestures are used to share interest with the other about some referent or provide him/her with information that might be useful. Imperative and declarative pointing, which both represent powerful means of establishing joint attention, have been extensively studied and compared:

hand shape and body posture were shown to differ according to the communicative function of the pointing gesture (Cochet et al., 2014), as well as the frequency of gaze alternation between the partner and the referent and the frequency of vocalizations (Cochet and Vauclair, 2010). These comparisons (see section "Pragmatics in HRI: Which Ingredients Are Necessary for Effective Interactions?" for more detailed results) thus highlight the strong relationship between the form of the gestures (in the broad sense, i.e., including visual and vocal behavior in addition to movement kinematics and hand shapes) and pragmatic features in children, even semantic ones in adults (Cochet and Vauclair, 2014).

To sum up, when two children are playing together or when a child is playing with an adult, they do so in the framework of joint action; they attend to a common situation and use multimodal communication to initiate, maintain, or respond to the interaction. These three different roles in the interaction can be assessed with the Early Social Communication Scales, in particular with the French version (Guidetti and Tourrette, 2017). In an evaluation situation, giving the child the opportunity to initiate the interaction is particularly crucial in atypical development, for example in children with ASD. The initiation of shared attention is a key ability in this context as it allows joint action coordination (Vesper et al., 2016) and has also significant consequences on the development of cognitive and emotional processes (Shteynberg, 2015). Whether this coordination relies on the representation and the understanding of the other's intentions or only on behavioral cues is a challenging question, as we do not have any direct access to the other's subjectivity. In the field of HRI, an objective that appears sufficiently ambitious for now, or at least the one we chose to focus on in the present review, is to design robots able to identify the observable changes in the human's behavior, in order to make the right inferences and thus the appropriate decisions in the interaction. This appears as an essential condition for a successful exchange between a robot and a human, which can depend on the joint outcome (has the common goal been reached?), but also on the way the interaction has been perceived by each individual, for example in terms of coordination between gaze and gesture and fluidity of movement (Hough and Schlangen, 2016). The richness of communication here lies indeed in the ability of each partner to integrate multiple communicative cues in a way that what will seem *natural* to the humans, i.e., that will be close to peer interaction in everyday life.

This appears as a complex ability and probably the most challenging one to replicate in HRI. In pursuit of this objective, we now need to further describe the concept of *appropriateness* and propose a frame to determine the relative importance and the relative complexity of the different behaviors observed during joint activities such as social play.

TO WHAT EXTENT CAN INTERACTIONS BE CHARACTERIZED AS *COMPLEX*?

Smith (2015) has argued that "development, like evolution and culture, is a process that creates complexity by accumulating change." This perspective applies to the development of social interactions, from the emergence of joint attention to coordinated and multimodal communication that enable joint action. Several attempts have been made in developmental robotics to explore the cognitive, social, and motivational dynamics of human interactions (Oudeyer, 2017); algorithmic and robotic models can then be used to study the developmental processes involved for instance in imitation (Demiris and Meltzoff, 2008) or language (Cangelosi et al., 2010). In this context, roboticists aim at designing systems allowing for self-organized and "progressive increase in the complexity" of the robot's behavior (Oudeyer et al., 2007).

To benefit further from their exchanges, developmentalists and roboticists may therefore need to frame the study of HRI by disambiguating the concept of complexity. Because "complicated systems will be best understood at the lowest possible level" (Smith, 2015), we aim to differentiate different levels of complexity depending on the nature of the elements to take into account for decision making. This analysis will allow us to go forward in the study of joint attention and joint action and define what is implied by the qualifying terms "complex" (or rich) and "appropriate" (or effective) when referring to interactions.

To this end, we used a categorization recently proposed in research on animal behavior, including human and non-human primates, to define the concept of complexity (Cochet and Byrne, 2015). Three dimensions have been described: motor precision, coordination, and anticipatory planning, which can relate to both individual and social activities. The authors argue that "the complexity of a given mechanism/behavior can be assessed by distinguishing which of these three dimensions are involved and to what degree," which may "clarify our understanding of animal behavior and cognition." Such analysis applied to joint attention and joint action, although there may be other ways of untangling the question of complexity, may here allow researchers to dissect the different factors involved in social interactions for each dimension, and thus help them assess the "manipulability" of these factors in HRI.

In order to make appropriate decisions in a collaborative task, i.e., decisions leading to the desired joint outcome and/or decisions that approach the characteristics of human interactions, the robot first needs to recognize specific patterns in his/her partners' behavior, without asking for agreement or information for all actions. The robot can for example rely on gaze direction, manual movements or body posture to identify the human's attentional and intentional states and thus define the most useful role it can play in the interaction. By way of illustration, if a human and a robot share the common goal of building a pile with four cubes in a definite order and putting a triangle at the top, each of them can perform different actions: they can grasp an object (a cube or a triangle) on the table, grasp an object on the pile, give an object to the partner, support the pile while the partner places a cube on it, etc. Other actions can emerge, for example if the pile collapses or if one agent does not pile the cubes in the correct order (Clodic et al., 2014). Individuals can then blame each other, or give each other some instructions. In addition to the perception of its own environment, the robot thus has to observe the activity of the human and take his/her perspective

(e.g., to determine whether an object is reachable for the other).

Motor precision is therefore necessary in this context to obtain flexible and human-aware shared plan execution (Devin and Alami, 2016), as it enables a selective shift of attention toward aspects of the environment that will become shared knowledge, which has also been described as the *accuracy* of shared attention states (Shteynberg, 2015). First, the emergence of joint attention requires to properly use gaze and/or pointing gesture to localize the object or event referred to. Verbal cues also demand particularly fine motor skills through speech articulators. Second, joint action necessitates some motor control to reach the expected outcome, hence the importance of evaluating beforehand human motor skills, especially during development, as well as the technical capabilities of the robot. Following on from the previous example, children's grasping skills in relation to the size of the cubes as well as the characteristics of robotic gripper to handle objects have to be finely described.

Moreover, recent experimental findings have shown that the execution of object-oriented actions is influenced by the social context such as the relative position of another person and the degree of familiarity with this person (Gianelli et al., 2013). Individuals perform for example more fluent reach-to-grasp movements, with lower acceleration peaks and longer reaction time when a partner is located close enough to be able to intervene on the same object than when he/she is farther away (Quesque et al., 2013). In addition, there is a significant relationship between the kinematic features of the actions and the actor's explicit social intention: movements have longer durations, higher elevations and longer reaction times when individuals place an object on a table for another person than when they place the object for a later personal use (Quesque and Coello, 2015). These variations, although they do not seem to be intentionally produced, have been suggested to facilitate the partner's detection of planned actions, thus enhancing potential interactions. These kinematic effects were indeed shown to influence the subsequent motor productions of an observer (Quesque et al., 2015). The motor characteristics of actions performed in a social context may therefore prime the perceiver to prepare and anticipate appropriate motor responses in the interaction.

The second dimension that can allow us to understand the complexity of joint activities pertains to the coordination between several communicative modalities and between interacting individuals. Whether joint action involves complementary or similar roles, it can be performed through several coordination processes, which can determine the *efficiency* of shared attention states (Shteynberg, 2015). Efficiency requires here a representational shift from the first-person singular to the first-person plural, as the partners attend to the same referent at the same time. The ability to monitor each other's attention and action, using behavioral cues such as gaze direction, facial expressions, gestures, and speech is essential for successful coordination. The intentional production of communicative signals, representing hints for one's partner, is also an efficient way of achieving joint outcomes.

Coordination is therefore necessary first at the individual level, so that the different communicative modalities such as gestures and gaze synchronize or follow one another in a natural order, i.e., acceptable with regard to human interaction patterns (see above). Each agent can then make decisions based on these signals, moderate their behavior accordingly and thus coordinate at the social level to reach a common objective. The ability to adjust one's behavior to others' actions during collaborative activities (including play) has been argued to "reach a higher degree of complexity when intentional and referential signals are directly addressed to specific individuals" (Cochet and Byrne, 2015). In order to build the pile of cubes, interacting partners can then for example point toward a specific cube or ask the other to wait before placing another cube.

In those cases, coordination processes can be enhanced by predicting the effects of each other's actions on joint outcomes and by distributing tasks effectively (Vesper et al., 2016). This ability involves the third dimension characterizing the question of complexity, namely the dimension of anticipatory planning (Cochet and Byrne, 2015). It requires to go beyond the immediate perception of the environment and represent the relationship between a sequence of actions and a precise goal. At the individual level, planning ability implies to mentally review an action sequence in anticipation of a future need (e.g., selecting a specific cube in a first room in order to build a pile of cubes in another room). At the social level, planning ability allows individuals to predict the other's behavior and adjust one's own sequence of actions, leading to a better coordination. Whether the ability to make such inferences necessitates to mentalize about others' inner states (e.g., beliefs and preferences) is still subject of debate, but again, this question may not be central in the context of joint attention and joint action between a robot and a human.

The above-described categorization can therefore provide a common ground between ethologists, psychologists, and roboticists that may clarify which dimensions need to be considered in an attempt to implement the characteristics of motor precision, coordination and anticipatory planning in human–robot joint activities (see **Table 1** for an overview). The objective is to approach the complexity (or richness) of human interactions and obtain appropriate (or effective) responses from robots with regard to these different dimensions.

PRAGMATICS IN HRI: WHICH INGREDIENTS ARE NECESSARY FOR EFFECTIVE INTERACTIONS?

The increasing complexity of communicative abilities (complexity that involves the three above-mentioned dimensions) in the course of human development leads to a rich potential of interactions. Children actively go through different stages allowing them to engage successfully in joint activities, i.e., to operate within their physical environment, coordinate with other people, plan their own behavior and anticipate their partners'. Intending to model, at least partially, human developmental pathway seems a fruitful way of designing robots that can effectively initiate and respond to communicative

TABLE 1 | Complexity in HRI: illustration of three dimensions at the individual and social levels (adapted from Cochet and Byrne, 2015).

	Individual	Social
(1) Motor precision	Joint attention: ability to properly use gaze or pointing to identify the object or event referred to. Joint action: human motor skills/technical capabilities of the robot to reach the expected outcome	Influence of the social context (e.g., relative position of the individuals, intention of the actor: moving an object for oneself or for another person) on the kinematics features of the actions performed
(2) Coordination	Coordination (including synchronization) between different modalities of one's communicative signal (gaze, gesture, vocalizations, etc.)	Ability to take into account the multimodal behavioral cues produced by a partner to adjust one's own behavior
(3) Anticipatory planning	Representation of a sequence of actions to anticipate a personal future need	Ability to predict the effects of the other's actions on joint outcomes to plan one's own behavior

situations. Such enterprise, although still recent, has given rise to a substantial amount of literature in robotics, especially from the 2000s, covering several sub-fields such as for example developmental and epigenetic robotics, cognitive systems and social robotics. Several journals, including both HRI experimental studies and computational modeling, focus entirely on these questions (e.g., *IEEE Transactions on Cognitive and Developmental Systems, Journal of Human-Robot Interaction, Journal of Social Robotics*), and numerous conferences also take place every year, whose proceedings are usually available online[1].

The data from developmental psychology described in the first section, coupled with the framework proposed in the second section to help researchers define complex and effective HRI, may contribute to this growing body of work. To this effect, it seems necessary (1) to consider the multimodality of interactions and (2) to adopt a pragmatic perspective to be based upon an accurate representation of human communicative behaviors. Indeed, children learn to communicate through joint activities with adults who combine various forms of expressions, serving various functions. In the course of development, children gradually integrate the dissociation between the form and the function of language – they become more and more flexible in understanding that a single form can serve different functions and reciprocally, that a single function can be expressed through several forms. Language is here regarded as more than a medium to convey an information, in agreement with a proposition that was developed in the speech act theory (Austin, 1962; Searle and Vanderveken, 1985). Language would be way of acting on the environment, of "doing things with words," independently of its structural properties. Initially aiming at describing the relationships between the forms and functions of linguistic utterances, this theory defines several speech acts, depending on whether one intends to assert, comment, warn, request, deplore, etc. This theory has later been adapted to non-verbal behavior (e.g., McNeill, 1998; Guidetti, 2002). The form still refers to the message structure, but applies to the whole body, including the posture, the structure of communicative gestures (kinematic features and hand shape), gaze and facial expressions. These non-verbal signals can be used in complementarity with speech or be used alone for example in the case of conventional gestures (see Guidetti, 2002). The function refers to the illocutionary force of the speech act (what one achieves by speaking), in other words here to the effect of these communicative acts in a specific context, thus giving some insight into the signaller's intention. Gestures, and especially the conventional gestures produced by children during the prelinguistic period, are thus regarded as genuine communicative acts, with a propositional content that can equal the one expressed by words. For instance, agreeing and refusing can be expressed gesturally by nodding or shaking one's head. The separate analysis of the forms and functions of communication, as well as the description of the different modalities involved during interactions, therefore provide a key framework to help define what capacities the robot should be equipped with to ensure efficient collaboration with humans.

In this perspective, Mavridis (2015) has proposed a list of "ten desiderata that human–robot systems should fulfill" to maximize communication effectiveness. One of the guiding lines relates to the importance of considering multiple speech acts, for both verbal and non-verbal communication, and not restrict the robot competencies to "motor command requests." In the same way as imperative gestures (see section "How Does Communication Develop in the Context of Social Play?") are generally understood and produced later than declarative gestures in human development (Camaioni et al., 2004), robotic systems initially aimed to assign the robot a servant role, with the human driving the interaction. Devising wider robots' pragmatic abilities is a first step toward the conception of human–robot shared plans. The robot may for example comment on the pile of cubes as it is being built (see example section "To What Extent Can Interactions Be Characterized as *complex*?") to support or correct the human's action, rather than just producing a motor response to the human request. The dimension of social coordination is thus added to that of motor precision (see **Table 1**).

Similarly, flexibility in HRI also requires "mixed initiative dialog" (Mavridis, 2015), so that the robot can both initiate and respond to the interaction. Integrating models based on human adaptation and probabilistic decision processes, Nikolaidis et al. (2017) have indeed shown that the performance of human–robot teams in collaborative tasks is improved when the robot guides the human toward an effective strategy, compared to the common approach of having the robot strictly adapting to the human. The human's trust in the robot was also facilitated by a greater symmetry in role distribution and adaptation between the robot and the human, which might in turn lead to greater acceptability of HRI.

Designing such "socially intelligent and cooperative robots" (Breazeal et al., 2004) requires specific temporal dynamics of the interaction, which represents a considerable challenge especially at a computational level. These dynamics convey social meanings

[1]For example, http://www.lucs.lu.se/epirob/

to such an extent that any delay in the interaction can sometimes question its effectiveness. Researchers here face a dilemma that seem to bring into opposition interaction complexity (which requires to take account of numerous parameters) and interaction timing. The implementation of fast timescales (on the order of 100 ms) is usually considered necessary for robots to integrate (i.e., detect, interpret, and predict) and react to social stimuli in a timely manner through interactions (Durantin et al., 2017). Researchers developing a storytelling robot interacting with children aged 4–5 years have confirmed the importance of temporal features in the pragmatics of interactions. Contingent responses from the robot, in relation to the attentional and social cues signaled by the children, were indeed found to facilitate engagement of the latter (Heath et al., 2017).

The variation in some characteristics of the robot's behaviors according to the action performed may also illustrate further the question of pragmatics in HRI, moving us one step closer toward human-like interactions. For example, the morphological differences that have been reported in young children between pointing and reaching (Cochet et al., 2014) could be applied to the robot. First, regarding body posture, we might expect robots to lean closer to a given object when they intend to grasp it than when they want to communicate about that object. Second, depending on the robot technical possibilities (e.g., two- or three-finger grippers, biomimetic anthropomorphic hands), differences in the form of manual gestures produced should be observed between imperative and declarative pointing. The former is typically characterized by whole-hand gestures (all the fingers are extended in the direction of the referent), while the latter is mostly associated with index-finger gestures (the index finger is extended toward the referent and the other fingers are curled inside the hand) (Cochet and Vauclair, 2010; Liszkowski and Tomasello, 2011). Hand shape is also influenced by precision constraints: imperative gestures are likely to shift from whole-hand pointing to index-finger pointing when the target is surrounded by distractors (Cochet et al., 2014), which can be the case when the robot has to identify a specific object among several (e.g., the human can ask the robot to give him/her the red cube). Here, the notion of iconicity, which plays a role in both oral and sign languages, may help researchers to precisely analyze the structure of gestures and better understand the interface between gestures and signs (Guidetti and Morgenstern, 2017). The importance of motor precision is here directly related to the dimensions of coordination and anticipatory planning, therefore providing a comprehensive framework to assess the complexity and effectiveness of HRI.

Moreover, the importance of implementing responsive social gaze in robots has previously been highlighted (e.g., Yoshikawa et al., 2006), but this response might also vary depending on the communicative function involved. To mirror child development, gaze alternation between the partner and the referent should indeed be more frequent in declarative situations than in imperative ones (Cochet and Vauclair, 2010). The coordination between gestures and gaze (see also section "How Does Communication Develop in the Context of Social Play?") is also an important factor, which can help the robot to estimate the state of goals, plans, and actions from human point of view, and allow the human to feel that he/she is involved in fluid interactions with the robot, both facilitating the emergence of joint outcomes. If a robot alternates its gaze between an object and its partner before initiating a pointing gesture, the human may for example interpret this behavior as the robot's willingness to take into account his/her attentional state before gesturing, thus favoring the exchange of information. Broadly speaking, coordinated gaze behavior could be considered as the most fundamental modality for effective HRI, or at least as a key prerequisite in collaborative tasks.

The consideration of facial expressions may also facilitate turn-taking dynamics and limit miscommunication, by allowing some inferences about the other's affective state. Integrating the emotional component into HRI gives each partner additional cues to decide what is the most appropriate response in a given situation. The development of methods for facial expression analysis raises several issues though (e.g., Kanade et al., 2000). Even if there have been some attempts to design facial expression mechanism in humanoid robots (e.g., Hashimoto et al., 2006; Gao et al., 2010), most of current robots' facial features are still far from the extremely rich motor possibilities of the human face. In parallel, the development of real time coding of emotional expressions seems to be an achievable goal (Bartlett et al., 2003), allowing robots to directly perceive some changes in the human facial expressions.

In addition to visual information, the auditory modality can also play a role in influencing robots' and humans' decisions and coordination processes. In children at around 2 years of age, vocalizations accompany more frequently declarative gestures than imperative ones (Cochet and Vauclair, 2010). More recently, the prosody of these vocalizations was shown to gradually match the function of pointing during the second year of life (Tiziana et al., 2017), allowing to differentiate imperative from declarative gestures (Grünloh and Liszkowski, 2015). Other features such as the positioning of the object and the attentional state of the partner have also been suggested to influence the rising and falling tones in the vocal productions simultaneous to gestures (Leroy et al., 2009). Prosody can therefore serve pragmatic purposes, and changes in pitch, intensity, or duration of speech or vocalizations can in this regard be considered as a full-fledged component of multimodal communication.

Beyond prosody, language content may be the most effective way for human–robot teams to coordinate. However, the design of robots with language comprehension and production abilities that could lead to fluid conversations with humans raises several issues. Verbal language requires indeed symbolic representations, which need to be connected not only to the robot's sensory system, but also to "mental models" of the world internalized within its cognitive system. Mavridis (2015) has highlighted here the question of "situated language and symbol grounding." For example, the relation between the verbal label "cube" uttered by the human and the physical cube that it refers to in front of the robot can be mediated through sensory data, but the use of conventional signs should allow the robots to go beyond the *here-and-now* and extend symbol grounding to abstract entities in addition to objects, people, or events. To implement architecture that can be compared to human interactions, this relation should

be bidirectional: the visual perception of a cube should activate the right symbol in the robot's cognitive system, leading to the production of the word "cube"; reciprocally, a request addressed to the robot to give the human the cube should create a precise representation, allowing the robot to identify the right object.

Moreover, the identification of emotion labels in the verbal modality could also contribute, in addition to the recognition of emotional facial expressions and acoustic properties of speech (see Breazeal, 2004 for a complete review on emotion systems in robots), to a better coordination between each partner of the interaction. The haptic modality, playing an important role in social interactions, is also regarded as a valuable medium for expressing emotion (Yohanan and MacLean, 2012). By developing motion capturing system and tactile sensors, the robot may use its human partner's positions and such "affective touch" to estimate human intentions (Miyashita et al., 2005). This modality, essential in human development, may be a particularly good candidate to study complexity of HRI, involving simultaneously motor precision, coordination and planning (see section "To What Extent Can Interactions Be Characterized as *complex*?").

Finally, in addition to the coordination dimension, the verbal dialog between a robot and a human would ideally imply purposeful speech and planning (Mavridis, 2015), in order to avoid fixed mapping between stimuli and responses. Anticipatory planning abilities, as described in Section "To What Extent Can Interactions Be Characterized as *complex*?", would enable the robot to make the most appropriate or efficient decisions in a given shared activity, in conjunction with its perspective-taking skills and the goal of the activity. If the robot can represent which information are needed by the human to perform a specific action (and therefore identify which information the human misses), it can decide to express a verbal request or comment on the situation, and/or plan a sequence of actions to coordinate with its partner.

This last example raises the question of intrinsic motivation in interactions: why is each partner engaged in this multimodal coordination, and to what extent does it influence the characteristics of the interaction? Studies in developmental robotics have shown that intrinsic motivation systems based on curiosity can directly impact learning skills and lead to autonomous mental development in robots (Oudeyer et al., 2007). Such mechanism is obviously involved in human development and in social play in particular: children discover and create new possibilities by exploring their physical and social environment. Through the development of social referencing, self-consciousness or cooperation, human social interactions may even sometimes constitute a motivated goal *per se* (Tomasello, 2009), which provides some perspectives to shape robots' intrinsic motivation with a "social reward" function.

We can see here that the relationships between theories in developmental psychology and robotics offer bidirectional benefits. To put it in a nutshell, some models in developmental robotics are based on psychological theories, which are then formalized and implemented in robots, while developmental robotics allows researchers in psychology to go further in the elaboration of their theories through thorough experimentations and hypothesis testing. This applies to a variety of questions addressed in this review, from the conditions that influence learning process during interactions (Boucenna et al., 2014) to the description of stages in language development (Morse and Cangelosi, 2017). Advances in developmental robotics may thus provide previous help in the analysis and implementation of the processes involved in interactions.

CONCLUSION AND PERSPECTIVES

The question at stake in the present work was to improve the effectiveness of human–robot interactions in collaborative tasks, first in terms of joint outcomes – has the task been completed? – but also with regard to the human's perception and interpretation of the interaction. Is the robot's behavior appropriate, i.e., acceptable, considering the frame of human communication? We argue here that the observation of the development and the structure of interactions between the child and the adult, especially in the context of social play, can help answer this question. To shape a shared common space between the human and the robot that could reflect the complexity of human interactions, we have also proposed to focus on three dimensions: motor precision, coordination, and anticipatory planning. The specific examples developed in Section "Pragmatics in HRI: Which Ingredients Are Necessary for Effective Interactions?" suggest that the more robots use human-like communicative modalities (e.g., facial expressions, gestures, and language) in respect to these three dimensions, the more they invite interactive behaviors that are natural to people. The interpretation of dealing with a social agent is strengthened, which facilitates in turn the interaction with robots. In this sense, and to paraphrase Cangelosi et al. (2010), the integration of action and language may constitute a roadmap to better frame and assess HRI from a developmental point of view and with a pragmatic perspective.

However, there are still numerous obstacles before achieving the level of details pictured in the present article, involving mainly technological challenges, given the motor and cognitive correlates of the above-mentioned behaviors. To put it bluntly, developmental psychologists cannot expect roboticists to implement in robots all the subtleties of multimodal communication that occur in human children. There may also be some conceptual difficulties as the attempts to approach human realism, aiming at maintaining the human's trust in the robot, can sometimes be confronted with an uneasy feeling of viewing and/or hearing a robot that looks imperfectly human. This *uncanny valley* effect (Mitchell et al., 2011; Mori, 1970, 2012), which was shown to emerge in middle childhood in relation to developing expectations about humans and machines (Brink et al., 2017), may complicate the design of socially interactive robots, both in terms of appearance and behavior. Empirical evidence for the uncanny valley seems nevertheless inconsistent or restricted to specific conditions (Kätsyri et al., 2015), with the definition of human-likeness mostly involving physical realism.

By contrast, anthropomorphic behavior (see Duffy, 2003), in addition to its facilitating role in the interaction with humans (see

above), also results in better and faster learning by the robots. For example, in a task in which they have to learn the meaning of words, the robots' performances are enhanced when they provide humans with social cues to communicate a learning preference, as these cues influence the tutoring of the human teacher (de Greeff and Belpaeme, 2015). We observe the same phenomena when human children start to learn new concepts: according to Bruner's constructivist theory, children need scaffolding from adults (or from children who have already acquired the concept) in the form of active support, which may represent at first a reduction in the choices a child might face. Such learning processes play obviously an important role in human development, and may also enable quick and effective application of robotic systems. Multi-level learning may indeed constitute a key line of research for HRI (Mavridis, 2015), which might again benefit from research in developmental psychology.

Reciprocally, the field of robotics provides interesting perspectives for psychologists, especially for research on atypical development. Atypical development might be a direct window on typical development and vice versa: "development is the key to understanding developmental disorders" (Karmiloff-Smith, 1998). Joint action and joint attention are for example usually impaired in children with ASD; the comparison with typical development has revealed different use of social gaze and often a lack of the declarative function, both for verbal and non-verbal communication. The exchanges between robotics and developmental psychology could help conceptualize the stages of joint attention in order to better understand how children develop joint attention and get through the whole sequence of declarative pointing. This will have an impact on elaborating intervention programs for children with neurodevelopmental disorders. Moreover, numerous intervention programs have recently been proposed showing the added value of therapy robot for the development of communication, play, or emotional skills (e.g., Robins et al., 2009; Huijnen et al., 2016).

In conclusion, the combination of insights and methods in robotics and developmental psychology allows researchers to conceive models of HRI in which the robots can come to develop motor, social, and cognitive skills. These models may benefit fundamental research on joint attention and joint action in typical development, but also early evaluation and intervention programs for atypical development (e.g., Dautenhahn, 2007). The continuation of these interdisciplinary discussions, which may possibly integrate some of the elements proposed in the present article, will undoubtedly lead to more and more solid HRI models in the next decades.

AUTHOR CONTRIBUTIONS

HC and MG devised the conceptual ideas presented in the article. HC drafted the manuscript. MG revised it critically and gave final approval of the version to be submitted.

FUNDING

This article is part of the project JointAction4HRI, funded by the French National Agency for Research (n°16-CE33-0017).

ACKNOWLEDGMENTS

Many ideas presented in this paper stem from fruitful discussions with R. Alami, A. Clodic, and E. Pacherie, all involved in the *Joint Action for Human-Robot Interaction* project funded by French National Agency for Research (Project No. 16-CE33-0017-01).

REFERENCES

Adamson, L. B., Bakeman, R., Suma, K., and Robins, D. L. (2017). An expanded view of joint attention: Skill, engagement and language in typical development and autism. *Child Dev.* doi: 10.1111/cdev.12973 [Epub ahead of print].

Aureli, T., Grazia, A., Cardone, D., and Merla, A. (2015). Behavioral and facial thermal variations in 3-to 4- month-old infants during the Still-Face Paradigm. *Front. Psychol.* 6:1586. doi: 10.3389/fpsyg.2015.01586

Austin, J. (1962). *How to do Things with Words*. Oxford: Clarendon Press.

Bartlett, M. S., Littlewort, G., Fasel, I., and Movellan, J. R. (2003). Real time face detection and expression recognition: development and application to human-computer interaction. *CVPR Workshop Comput. Vis. Patt. Recogn. Hum. Comput. Interact.* 5, 53–58.

Bates, E., Camaioni, L., and Volterra, V. (1975). The acquisition of performatives prior to speech. *Merrill Palmer Q.* 21, 205–226.

Beier, J. S., and Spelke, E. S. (2012). Infants' developing understanding of social gaze. *Child. Dev.* 83, 486–496. doi: 10.1111/j.1467-8624.2011.01702.x

Beuker, K. T., Rommelse, N. J., Donders, R., and Buitelaar, J. K. (2013). Development of early communication skills in the first two years of life. *Infant Behav. Dev.* 36, 71–83. doi: 10.1016/j.infbeh.2012.11.001

Boucenna, S., Anzalone, S., Tilmont, E., Cohen, D., and Chetouani, M. (2014). Learning of social signatures through imitation game between a robot and a human partner. *IEEE Trans. Auton. Mental Dev.* 6, 213–225. doi: 10.1109/TAMD.2014.2319861

Bourjade, M. (2017). "Social attention," in *The International Encyclopedia of Primatology*, ed. A. Fuentes (New York, NY: John Wiley and Sons).

Breazeal, C. (2004). Function meets style: Insights from emotion theory applied to HRI. *IEEE Trans. Syst. Man Cybern.* 34, 187–194. doi: 10.1109/TSMCC.2004.826270

Breazeal, C., Brooks, A., Gray, J., Hoffman, G., Kidd, C., Lee, H., et al. (2004). Humanoid robots as cooperative partners for people. *Int. J. Hum. Rob.* 1, 1–34. doi: 10.3389/fnbot.2012.00003

Brink, K. A., Gray, K., and Wellman, H. M. (2017). Creepiness creeps in: Uncanny valley feelings are acquired in childhood. *Child Dev.* doi: 10.1111/cdev.12999 [Epub ahead of print].

Bruner, J. S. (1973). Organization of early skilled action. *Child Dev.* 44, 1–11. doi: 10.2307/1127671

Butcher, C., and Goldin-Meadow, S. (2000). "Gesture and the transition from one- to two-word speech: when hand and mouth come together," in *Language and Gesture*, ed. D. McNeill (Cambridge: Cambridge University Press), 235–257.

Byrne, R. W. (2015). The what as well as the why of animal fun. *Curr. Biol.* 25, R2–R4. doi: 10.1016/j.cub.2014.09.008

Camaioni, L., Perucchini, P., Bellagamba, F., and Colonnesi, C. (2004). The role of declarative pointing in developing a theory of mind. *Infancy* 5, 291–308. doi: 10.1207/s15327078in0503_3

Cangelosi, A., Metta, G., Sagerer, G., Nolfi, S., Nehaniv, C., Fischer, K., et al. (2010). Integration of action and language knowledge: a roadmap for developmental robotics. *IEEE Trans. Autonom. Mental Dev.* 2, 167–195. doi: 10.1109/TAMD.2010.2053034

Carpenter, M., Nagell, K., and Tomasello, M. (1998). Social cognition, joint attention, and communicative competence from 9 to 15 months of age. *Monogr. Soc. Res. Child Dev.* 63:255. doi: 10.2307/1166214

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., and Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cogn. Dev.* 15, 481–498. doi: 10.1016/S0885-2014(01)00037-5

Clodic, A., Alami, R., and Chatila, R. (2014). Key elements for human-robot joint action. *Front. Artif. Intell. Appl.* 273, 23–33.

Cochet, H., and Byrne, R. W. (2015). Complexity in animal behaviour: towards common ground. *Acta Ethol.* 18, 237–241. doi: 10.1007/s10211-014-0205-5

Cochet, H., and Byrne, R. W. (2016). Communication in the second and third year of life: relationships between nonverbal social skills and language. *Infant Behav. Dev.* 44, 189–198. doi: 10.1016/j.infbeh.2016.07.003

Cochet, H., Jover, M., Oger, L., and Vauclair, J. (2014). Morphological differences between imperative and declarative pointing: hand shape, arm extension and body posture. *J. Motor Behav.* 46, 223–232. doi: 10.1080/00222895.2014.889066

Cochet, H., and Vauclair, J. (2010). Pointing gestures produced by toddlers from 15 to 30 months: different functions, hand shapes and laterality patterns. *Infant Behav. Dev.* 33, 432–442. doi: 10.1016/j.infbeh.2010.04.009

Cochet, H., and Vauclair, J. (2014). Deictic gestures and symbolic gestures produced by adults in an experimental context: hand shapes and hand preferences. *Laterality* 19, 278–301. doi: 10.1080/1357650X.2013.804079

Colonnesi, C., Stams, G., Koster, I., and Noom, M. J. (2010). The relation between pointing and language development: a meta-analysis. *Dev. Rev.* 30, 352–366. doi: 10.1016/j.dr.2010.10.001

Corkum, V., and Moore, C. (1998). Origins of joint visual attention in infants. *Dev. Psychol.* 34, 28–38. doi: 10.1037/0012-1649.34.1.28

Dautenhahn, K. (2007). Socially intelligent robots: dimensions of human–robot interaction. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 362, 679–704. doi: 10.1098/rstb.2006.2004

de Greeff, J., and Belpaeme, T. (2015). Why robots should be social: enhancing machine learning through social human-robot interaction. *PLoS One* 10:e0138061. doi: 10.1371/journal.pone.0138061

Demiris, Y., and Meltzoff, A. (2008). The robot in the crib: a developmental analysis of imitation skills in infants and robots. *Inf. Child Dev.* 17, 43–53. doi: 10.1002/icd.543

D'Entremont, B., and Seamans, E. (2007). Do infants need social cognition to act socially? An alternative look at infant pointing. *Child Dev.* 78, 723–728. doi: 10.1111/j.1467-8624.2007.01026.x

Devin, S., and Alami, R. (2016). "An implemented theory of mind to improve human-robot shared plans execution," in *Proceedings of the 11th ACM/IEEE International Conference HRI*, New York, NY, 319–326.

Duffy, B. R. (2003). Anthropomorphism and the social robot. *Rob. Auton. Syst.* 42, 177–190. doi: 10.1016/S0921-8890(02)00374-3

Durantin, G., Heath, S., and Wiles, J. (2017). Social moments: a perspective on interaction for social robotics. *Front. Robot. AI* 4:24. doi: 10.3389/frobt.2017.00024

Emery, N. J. (2000). The eyes have it: The neuroethology, function and evolution of social gaze. *Neurosci. Biobehav. Rev.* 24, 581–604. doi: 10.1016/S0149-7634(00)00025-7

Farroni, T., Menon, E., Rigato, S., and Johnson, M. H. (2007). The perception of facial expressions in newborns. *Eur. J. Dev. Psychol.* 4, 2–13. doi: 10.1080/17405620601046832

Franco, F., and Butterworth, G. (1996). Pointing and social awareness: declaring and requesting in the second year. *J. Child Lang.* 23, 307–336. doi: 10.1017/S0305000900008813

Gao, J., Huang, Q., Yu, Z., Chen, X., Xu, W., and Zhang, Y. (2010). Design of the facial expression mechanism for humanoid robots. ROMANSY 18 Rob. *Design Dynam. Control* 524, 433–440.

Garvey, C. (1990). *Play*. Cambridge, MA: Harvard University Press.

Gianelli, C., Scorolli, C., and Borghi, A. M. (2013). Acting in perspective: the role of body and language as social tools. *Psychol. Res.* 77, 40–52. doi: 10.1007/s00426-011-0401-0

Grünloh, T., and Liszkowski, U. (2015). Prelinguistic vocalizations distinguishing pointing acts. *J. Child Lang.* 42, 1312–1336. doi: 10.1017/S0305000914000816

Guidetti, M. (2002). The emergence of pragmatics: forms and functions of conventional gestures in young French children. *First Lang.* 22, 265–285. doi: 10.1177/014272370202206603

Guidetti, M. (2005). Yes or no? How do young children combine gestures and words to agree and refuse. *J. Child Lang.* 32, 911–924. doi: 10.1017/S0305000905007038

Guidetti, M., and Morgenstern, A. (2017). The gesture sign interface in language acquisition. *Lang. Interact. Acquis.* 8, 1–12. doi: 10.1075/lia.8.1.01gui

Guidetti, M., and Tourrette, C. (2017). *Echelle d'Evaluation de la Communication Sociale Précoce*, 1st Edn. Firenze: Giunti OS.

Hashimoto, T., Hitramatsu, S., Tsuji, T., and Kobayashi, H. (2006). Development of the face robot Saya for rich facial expressions. *Proc. IEEE SICE-ICASE Int. Joint Conf.* 5423–5428. doi: 10.1109/SICE.2006.315537

Heath, S., Durantin, G., Boden, M., Hensby, K., Taufatofua, J., Olsson, O., et al. (2017). Spatiotemporal aspects of engagement during dialogic storytelling child-robot interaction. *Front. Robot. AI* 4:27. doi: 10.3389/frobt.2017.00027

Hough, J., and Schlangen, D. (2016). "Investigating fluidity for human-robot interaction with real-time, real-world grounding strategies," in *Proceedings of the 17th Annual SIGdial Meeting on Discourse and Dialogue*, Los Angeles, CA, 288–298. doi: 10.18653/v1/W16-3637

Huijnen, C. A. G. J., Lexis, M. A. S., and de Witte, L. P. (2016). Matching robot KASPAR to autism spectrum disorder (ASD): therapy and educational goals. *Int. J. Soc. Rob.* 8, 445–455. doi: 10.1007/s12369-016-0369-4

Kanade, T., Cohn, J., and Tian, Y. (2000). "Comprehensive database for facial expression analysis," in *Proceedings of the 4th International Conference Face Gesture Recognition*, London, 46–53.

Karmiloff-Smith, A. (1998). Development itself is the key to understanding developmental disorders. *Trends Cogn. Sci.* 2, 389–398. doi: 10.1016/S1364-6613(98)01230-3

Kätsyri, J., Förger, K., Mäkäräinen, M., and Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Front. Psychol.* 6:390. doi: 10.3389/fpsyg.2015.00390

Kendon, A. (2004). *Gesture: Visible Action as Utterance*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511807572

Lamaury, A., Cochet, H., and Bourjade, M. (2017). Acquisition of joint attention by olive baboons gesturing toward humans. *Anim. Cogn.* doi: 10.1007/s10071-017-1111-9 [Epub ahead of print].

Lemaignan, S., Warnier, M., Clodic, A., and Alami, R. (2017). Artificial cognition for social human-robot interaction: an implementation. *Artif. Intell.* 247, 45–69. doi: 10.1016/j.artint.2016.07.002

Leroy, M., Mathiot, E., and Morgenstern, A. (2009). "Pointing gestures, vocalizations and gaze: two case studies," in *Studies in Language and Cognition*, eds J. Zlatev, M. Andren, M. J. Falck, and C. Lundmark (Newcastle upon Tyne: Cambridge Scholars Publishing), 261–275.

Liebal, K., Carpenter, M., and Tomasello, M. (2013). Young children's understanding of cultural common ground. *Br. J. Dev. Psychol.* 31, 88–96. doi: 10.1111/j.2044-835X.2012.02080.x

Liszkowski, U., and Tomasello, M. (2011). Individual differences in social, cognitive, and morphological aspects of infant pointing. *Cogn. Dev.* 26, 16–29. doi: 10.1016/j.cogdev.2010.10.001

Lœvenbruck, H., Vilain, C., and Dohen, M. (2008). From gestural pointing to vocal pointing in the brain. *Rev. Fr. Ling. Appl.* 13, 23–33.

Mavridis, N. (2015). A review of verbal and non-verbal human-robot interactive communication. *J. Rob. Auton. Syst.* 63, 22–35. doi: 10.1016/j.robot.2014.09.031

McNeill, D. (1998). "Speech and gesture integration," in *The Nature and Functions of Gesture in Children's Communication*, eds J. M. Iverson and S. Goldin-Meadow (San Francisco, CA: Jossey-Bass).

McNeill, D. (2000). *Language and Gesture: Window into Thought and Action*. Cambridge: Cambridge University Press. doi: 10.1017/CBO9780511620850

Milward, S. J., Kita, S., and Apperly, I. A. (2017). Individual differences in children's corepresentation of self and other in joint action. *Child Dev.* 8, 964–978. doi: 10.1111/cdev.12693

Mitchell, W. J., Szerszen, K. A., Lu, A. S., Schermerhorn, P. W., Scheutz, M., and MacDorman, K. F. (2011). A mismatch in the human realism of face and voice produces an uncanny valley. *I-Perception* 2, 10–12. doi: 10.1068/i0415

Miyashita, T., Tajika, T., Ishiguro, H., Kogure, K., and Hagita, N. (2005). "Haptic communication between humans and robots," in *Proceedings of the 12th International Symposium Robot Research*, Berlin, 525–536.

Mori, M. (1970). Bukimi no tani [The uncanny valley]. *Energy* 7, 33–35.

Mori, M. (2012). The uncanny valley. *IEEE Robot. Autom. Mag.* 19, 98–100. doi: 10.1109/Mra.2012.2192811

Morse, A. F., and Cangelosi, A. (2017). Why are there developmental stages in language learning? A developmental robotics model of language development. *Cogn. Sci.* 41, 32–51. doi: 10.1111/cogs.12390

Nikolaidis, S., Hsu, D., and Srinivasa, S. (2017). Human-robot mutual adaptation in collaborative tasks: models and experiments. *Int. J. Robot. Res.* 36, 618–634. doi: 10.1177/0278364917690593

Oudeyer, P. Y. (2017). What do we learn about development from baby robots? *Wiley Interdiscip. Rev. Cogn. Sci.* 8:e1395.

Oudeyer, P. Y., Kaplan, F., and Hafner, V. V. (2007). Intrinsic motivation systems for autonomous mental development. *IEEE Trans. Evol. Comput.* 11, 265–286. doi: 10.1109/TEVC.2006.890271

Pons, F., Harris, P. L., and de Rosnay, M. (2004). Emotion comprehension between 3 and 11 years: developmental periods and hierarchical organization. *Eur. J. Dev. Psychol.* 1, 127–152. doi: 10.1080/17405620344000022

Quesque, F., and Coello, Y. (2015). Perceiving what you intend to do from what you do: evidence for embodiment in social interactions. *Socioaffect. Neurosci. Psychol.* 5:28602. doi: 10.3402/snp.v5.28602

Quesque, F., Delevoye-Turrell, Y., and Coello, Y. (2015). Facilitation effect of observed motor deviants in a cooperative motor task: evidence for direct perception of social intention in action. *Q. J. Exp. Psychol.* 69, 1451–1463. doi: 10.1080/17470218.2015.1083596

Quesque, F., Lewkowicz, D., Delevoye-Turrell, Y. N., and Coello, Y. (2013). Effects of social intention on movement kinematics in cooperative actions. *Front. Neurorobot.* 7:14. doi: 10.3389/fnbot.2013.00014

Robins, B., Dautenhahn, K., and Dickerson, P. (2009). "From isolation to communication: a case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot," in *Proceedings of the 2nd International Conference Advance Computing-Human Interaction*, New York, NY.

Scassellati, B. (2000). "How robotics and developmental psychology complement each other," in *Proceedings of the NSF/DARPA Workshop Development Learning Michigan State University*, Lansing, MI.

Searle, J. R., and Vanderveken, D. (1985). *Foundations of Illocutionary Logic*. Cambridge: Cambridge University Press.

Shteynberg, G. (2015). Shared attention. *Perspect. Psychol. Sci.* 5, 579–590. doi: 10.1177/1745691615589104

Smith, L. B. (2015). "Foreword," in *Developmental robotics – From babies to robots. Cambridge*, eds A. Cangelosi and M. Schlesinger (London: The MIT Press).

Tiziana, A., Spinelli, M., Fasolo, M., Garito, M. C., Perucchini, P., and D'Odorico, L. (2017). The pointing-vocal coupling progression in the first half of the second year of life. *Infancy* 22, 801–818. doi: 10.1111/infa.12181

Tomasello, M. (2009). *Why We Cooperate*. Cambridge, MA: MIT Press.

Tomasello, M., and Carpenter, M. (2007). Shared intentionality. *Dev. Sci.* 10, 121–125. doi: 10.1111/j.1467-7687.2007.00573.x

Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi: 10.1017/S0140525X05000129

Tronick, E. Z., Als, H., Adamson, L., Wise, S., and Brazelton, T. B. (1978). The infant's response to entrapment between contradictory messages in face-to-face interaction. *J. Am. Acad. Child Psychol.* 17, 1–13. doi: 10.1016/S0002-7138(09)62273-1

Vesper, C., Schmitz, L., Safra, L., Sebanz, N., and Knoblich, G. (2016). The role of shared visual information for joint action coordination. *Cognition* 153, 118–123. doi: 10.1016/j.cognition.2016.05.002

Vygotsky, L. S. (1999). "Tool and sign in the development of the child," in *The Collected Works of L. S. Vygotsky*, Vol. 6, ed. R. W. Rieber (New York, NY: Kluwer Academic/Plenum), 3–68.

Warneken, F., Chen, F., and Tomasello, M. (2006). Cooperative activities in young children and chimpanzees. *Child Dev.* 77, 640–663. doi: 10.1111/j.1467-8624.2006.00895.x

Yohanan, S., and MacLean, K. E. (2012). The role of affective touch in human-robot interaction: Human intent and expectations in touching the haptic creature. *Int. J. Soc. Robot.* 4, 163–180. doi: 10.1007/s12369-011-0126-7

Yoshikawa, Y., Shinozawa, K., Ishiguro, H., Hagita, N., and Miyamoto, T. (2006). Responsive robot gaze to interaction partner. *Conf. Robot. Sci. Syst.* 2, 287–293.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Cochet and Guidetti. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.