

2

Some boundary conditions on embodied agents sharing a common world

John Barresi

2.1 Introduction

What are the conditions that make it possible that you and I share the same world? This is the issue on which I wish to focus this chapter. There are a number of such conditions, some of which I would call boundary conditions. For instance, if you and I were in distant locations in space and time, it would be difficult for us to share a *common* world, even though conceptually, we might imagine that such a thing could exist for us. If you are outside of my light-cone, according to relativity theory, we cannot share a common world in any practical way. In order to share such a world we would need, at minimum, to have *personal* worlds that are within light-cone proximity to each other.

This is an example of what I mean by a boundary condition on embodied agents sharing a common world. It is a necessary condition on any form of world sharing between embodied agents that they are in light-cone relationship to each other. This is a necessary condition because without it there could be no overlap in personal worlds between one embodied agent and another; and, without overlapping personal worlds, there could be no *common world* for the agents to share.

Why am I asking the question of what are the boundary conditions for world sharing for embodied agents? I'm interested in the possibility of developing some common language to describe social relations between embodied agents of all kinds, from different species of animals, to humans, to robots. How can we describe their relations with each other in such a way that the language we use is general enough to cover all cases but affords differentiation? Searching for such a language may provide insight into basic principles of social life. Among the terms and their relations to each other that I will be discussing are *communication*, *causation*, *embodied agent*, *information*, *purpose*, *signal*, *embodied communication*, *intentional relation*, *personal world*, *common world*, and *common ground*. The hope is that by looking at these terms in this context, what might otherwise be lost in diverse terminological investigations of separate terms can be brought together within single overarching framework—one that I hope will be useful for reflecting on the diverse investigations undertaken in different chapters of this book.

2.2 Communication, causation, information, and embodied agents

In this attempt to determine basic principles of social life, the first term, which requires interpretation, is *communication*, for it is through communication, in the sense employed here that contact is made between personal worlds, making shared or common worlds possible. By the notion of communication between embodied agents considered as a boundary condition—or minimally necessary condition—for bringing about a world shared between them, I do not wish to presuppose that the communication between the agents is intentional. Nor does it necessarily involve transmitting information through signals with specific meanings. All that is required for *communication* of a minimal sort is some sort of causal relation between the agents, so that the behavior of at least one of the agents can be affected by information received from the other agent. This condition is minimal in two senses: first, the relationship can be asymmetric. Only one agent may be in the condition of communicative relation with the other. Second, there is no sent message or meaningful signal in the communication, just some sort of *causal process which transmits information* from the sender that can be *interpreted* in some manner by the receiver and modify the receiver's behavior.

There is another condition that is also necessary for minimal use of the notion of communication as I intend it here—that is that we have at least two *embodied agents* between whom the communication can occur. There are many causal relations between objects within light-cone distance to each other, but not all of these involve potential communications between embodied agents. Right now I'm capable of seeing in the night sky a supernova that may have happened millions of years ago from the point of view of my light cone. While this event may be part of the world that you and I can potentially share through communication with each other, it doesn't make sense to suggest that the supernova communicates with either of us. This is because a supernova is not itself an embodied agent. It is certainly an object that has causal consequences at far distances in space-time and can thereby transmit information about itself to an embodied agent; but this is not sufficient.

By *embodied agent*, I mean any sort of organic or robotic agent, and perhaps even virtual-quasi-embodied agents like Max (see Kopp *et al.*, this volume) that has purposes and actions of its own. Leaving aside boundary cases like Max, by embodied agents that are capable of communication I mean functionally unified material objects that transmit information about themselves as purposive agents through their actions and are capable of receiving and interpreting information about other purposive agents. And by *information* about another agent I mean any sort of *causal relation* that can be *interpreted* by one embodied agent as communicative about another embodied agent, in particular, about the latter's purposes and *personal world* though distant in space-time. This interpreted information from the sending agent can then be used by the receiving agent to modify its actions in order to achieve its own purposes in light of the activities and perceived purposes of the other agent. A common world is constructed through the modification of one personal world from communications about another personal world.

Quite a bit is packed into the notion of embodied agents in communicative relations or, more briefly, in *embodied communication*. However, the key here in this minimal

condition for communication is that embodied agents have a way of *interpreting* certain causal relations as *informational* relations between themselves and some other embodied agent. In order for embodied communication of this very general sort to occur at all, the receiver must be an embodied (or quasi-embodied) agent that can interpret causes issuing from another embodied agent as something about that agent's *personal world*. Otherwise there is no potential for a *common world* to be shared between the two agents—even an asymmetrical one, where only one agent has knowledge of that common world.

2.3 The personal and common worlds of embodied agents

2.3.1 Embodied agents, purposes, and personal worlds

I've used the terms *personal* and *common worlds* of embodied agents and these terms need to be examined. The more fundamental term is *personal world*. I am supposing that any embodied agent has a personal world. Consider the difference between a computer and a robot. Both a computer and a robot may be said to have programs that interpret causal relations with other objects as informational relations. Hence, it might appear that both are embodied agents in the sense I use the term. However, I take it that there is a crucial difference between our conception of a robot and our conception of a computer. Whereas robots are conceived to have purposes or goals of their own, computers are used to achieve the purposes or goals of agents who have purposes or goals of their own. For instance, a computer in a robot may take inputs from what we would call the sensory surfaces of the robot and transform them into information about objects, and, possibly, about other agents, in the robot's environment. However, though the computer may serve this function, it is a function embedded in the architecture of the robot—a function serving *purposes* of the robot, not purposes of its own. The same relationship holds between a brain in an organism and the organism. While the brain may process information received on the sensory surfaces of the organism, it is for the purposes of the organism, not for itself that it processes this information. When we think of embodied agents engaged in purposive activity, we have in mind *purposes* that an object can reasonably be said to have *on its own* and it could possibly attribute to itself, if it had the appropriate self-reflective apparatus (cf. Dennett 1987, for an alternative analysis, and Barresi 1999, for a fuller discussion of this issue). Such purposes, I would suggest, provide the constructive or constitutive basis of its own world—its *personal world*.

The key point here is that embodied agents—whether they are constructed following functions or purposes associated with evolution or through human technology—have purposes or goals, which *seem* to be their own, or are *best attributed* to them rather than to, or in addition to, the causes or agents that led to the object's construction. It is these purposes that are the constitutive basis of their *personal worlds*.

2.3.2 Intentional relations and personal worlds

How do we get from purposes attributed to an embodied agent to its personal world? In previous publications (e.g. Barresi and Moore 1996, 2008; Barresi 2007), Chris Moore and I have used the notion of *intentional relations* to capture an embodied agent's purposive

relations with other objects in its environment, or reflexively to itself as an object. I suggest here that it is out of these relations that the agent's *personal world* is constructed. Moore and I take *intentional relations* to be real causal relations between embodied agents and objects, but causal relations that involve purposes, such as *actions*, whether consciously intentional or not, that are directed at goals. We also consider *motivations* such as the need for food or sex, and *emotions* like anger and fear that lead to goal-directed actions as intentional relations. They are intentional relations because they are object-directed, in the sense that they can be interpreted as specifying an object in some general way that can be taken as *intentional* by being *about* some object or set of objects (see Barresi 2007, on intentionality and intentional relations). Finally, we include *epistemic* states such as seeing an object, or thinking about an object, as intentional relations, again because they have the property of aboutness, and ultimately relate to purposive actions of the embodied agent. In our view most intentional relations involve causal relations between an embodied agent and real existent objects currently in the agent's immediate environment, or in recent past or anticipated future environments; however some relate to counterfactual objects or situations.

We believe that the notion of intentionality in the philosophical sense following Brentano (1874), which tends to focus on counterfactual or representational relations between minds and what are called intended objects, is too complex a notion to invoke as a starting point in trying to understand how embodied agents relate to their worlds (Barresi 2007). Furthermore, this representational interpretation of intentionality has led to "disembodied" approaches to mind like "methodological solipsism" and the dismissal of "naturalistic psychology" grounded in causal relations between organisms and their environments (Fodor 1980). By contrast, the notion of intentional relation tries to restate a naturalistic psychology for embodied agents. We believe that it is out of causally grounded intentional relations, that more complex reflective and representational forms of conceptual intentional relations emerge in human development. And it is these more representational kinds of relations that provide the basis for the abstract notion of intentionality (Barresi 2007). Thus, it is one's embodied actual causal intentional relations with objects in one's immediate environment that ground more representational and counterfactual intentional relations associated with philosophical intentionality. And one's *personal world* is the world constructed out of this ensemble of causally grounded, as well as sometimes counterfactually constructed, intentional relations. It is a perspectival or *interpreted* world, one that expresses both purposes and ways of processing information that are relational between the embodied agent and the objects of its intentional relations.

Although the concept of personal world may be ill defined when it comes to artificially constructed objects like robots, the concept is reasonably well defined for biological organisms. All animals with central nervous systems can be readily attributed personal worlds. These animals engage in self-sustaining purposive activities that maintain their existence through acquiring energy resources while avoiding damage to their self-integrity and also engage in activities necessary for reproductive success (cf. Jordan, this volume). While the animals themselves are not directly motivated to acquire energy and

maintain self-integrity *per se*, or to achieve reproductive success, *per se*, they are constructed in a manner so that their more immediate purposes collectively tend to achieve these outcomes. However, each animal is constructed to achieve these evolutionary ends in different ways, which involve different energy sources, different means for self-preservation, and different roads to reproductive success. These different means involve a vast complex of contingently determined, intentional relations to objects in their environment, which collectively define their personal worlds. What makes them *personal* worlds is that the same objects can have different interpretations or *meanings* for different organisms, even different organisms in the same species. Indeed, the same object can change in meaning or intentional relations to the same organism over time so play different roles in the life history of the organism. Thus, personal worlds are both historical and semantic entities, since the embodied agent and the objects of its intentional relations both change through time, yet are intimately related to each other in ways that are at least partially independent of other embodied agents and their relations to the same objects.

2.3.3 Common worlds

We can turn now to the notion of *common worlds*. Common worlds emerge from the overlap and interaction of personal worlds. Each embodied agent has a personal world bound by its space–time cone of potential activity. It can only interact with objects in this cone. When two agents share regions in their space–time cones, there is a potential for interaction between them and the sharing of a common world. However, there are degrees or kinds of common worlds, and the variety of kinds of common worlds is of particular theoretical interest. Our main concern will be with common worlds that become shared as a result of communications between individuals about their personal worlds. In communication-based common worlds the personal worlds of agents come to overlap and form common worlds, through information about one or more of the agents' personal worlds being transmitted to other agents, and being incorporated into the receiving agents' personal worlds. This transmission can be accidental or purposive. When purposive, information transmission can involve *signals* (Tinbergen 1964; Marler 1965; Seyfarth and Cheney 2003), whose meaning indicates some aspect of the sender's personal world, which is understood by the receiver of the signal.

So far we have been considering here communication-based common worlds, where at least one agent and part of its personal world appears in some form in the personal world of the other agent. However, there is another, simpler way in which two embodied agents might be said to enjoy a common world. This could happen if both agents include some other object in their personal worlds that are in both of them. For instance, if both agents hear the sound of another object distant from both of them, they might have similar epistemic relations with that object. However, this kind of potential overlap in personal worlds would not involve any communication between embodied agents, thus would not constitute a shared common world in the sense used here. Unless eventually shared in a social way, for instance by one or both agents responding to the stimulus and having that response noticed by the other agent, or through the interlinking of personal worlds of multiple agents in a socially constructed world, communality of intentional relations of

this sort would not be of interest for this chapter. In later sections we will consider examples of common worlds that get formed in these more complex manners.

2.4 Embodied communication and signaling in competitive situations

At this point I could attempt a discussion of communication and common worlds that might be constructed between embodied agents whose personal worlds are radically different, for instance between alien creatures living at enormous distances in space–time from each other, or of embodied agents in the same physical world but so different in size that they hardly appear in each other’s personal worlds (Wilson 1975). Though conceptually important with respect to boundary conditions on common worlds, such discussions would not help us understand the relations between personal and common worlds in their simpler forms. Nor would it help at this point to consider personal and common worlds of robots, for it is to such embodied agents that we hope that the present analysis of embodied agents can be generalized. So let us focus instead on common worlds that exist between different species of organic agents who live close to each other in space–time here on earth—effectively in the same space–time cone—and are comparable in size. I will also drop pure asymmetric relations and begin by considering the case of predator–prey relations where the predator and prey are mutually aware of each other some of the time, though direct awareness can be asymmetric at other times, for instance just preceding such a mutual encounter. However, in the usual situation, when predator and prey are not in close personal proximity, they are only in a preparatory state, ready to pick up cues of each other, in a categorical sense, rather than involved in direct perception of particular others.

In the predator–prey relation each has its own personal world, mostly independent of the other (e.g. Schaller 1976). Part of the predator’s world includes prey as food, which generates a variety of intentional relations aimed at obtaining this energy resource. The prey has a quite different set of intentional relations to its energy resource. So, they have quite different personal worlds in this respect. But they share one thing in common, an interest in themselves and each other. In actual encounters, both the predator and prey attend to the prey and its position relative to the predator. Whereas the predator wishes to close the gap in its position relative to the prey, the prey wants to increase this gap. Thus both are highly sensitive to their relative positions to each other and to opportunities to increase or decrease the distance between them. These opportunities include the part of the environment that is apparently open to either of the animals in the near future, which they perceive primarily through visual attention to the environment in the direction of common motion, as well as the relative motions, and apparent intentions to change direction of either of the agents. Each of the agents assesses the activities of the other agent, and what it can do in the environmental situation to outdo the other agent in order to achieve its own personal goal, either escape or capture.

Predator–prey relations are a paradigm case of competitive relations among organisms, representative of “nature, red in tooth and claw.” However, from the point of view of

personal and common worlds, what is striking here is that although their personal worlds are generally radically different from each other when they are not engaged in interaction, they also have periods of interactive activity in which there is a common world between them that is incredibly articulated. In the case of large animals, like lions and antelopes, it is as if the predator and prey are dancing together at high speed in a wild environment, each agent highly sensitive to the motions of the other agent and responsive not only to past and present motions, but also to anticipated future motions. Although, the behavioral analysis by one of these agents of the other's activity does not—in the usual case—require an understanding of mental activities of the other agent, there is, nevertheless, a sophisticated model of the capacities and likely motions of the other agent that must govern this dance of life and death. This model appears to be one that involves entrainment to each other's motions of the physical oscillatory processes that govern their own individual motions (Rizzolatti 2005; Wilson and Knoblich 2005; Wolpert *et al.* 2003 Richardson *et al.* this volume). Because survival depends on success at escape or capture, depending on the agent, they have been adapted to predict the other's behavior with maximal accuracy and minimal delay, and to refine their own action intentional relations, in order to respond to the other in a personally successful manner.

In the coadapted dance of relative motions of predator and prey there is embodied communication of a rich sort and a kind of dialogue going on. Each animal can interpret information about the other animal's movements as messages, mainly "true" about the direction and speed of its motions, now and into the future. But each animal is also engages in false communications that are more purposive in their intent, as each tries to send misleading or "false" information about the actual future direction of its motions to the other. Neither the true information transmitted to the other animal, nor the false information, can be viewed as consciously intentional communications from the point of view of either actor, but as species-specific behavioral tendencies that in effect serve as communications. Each animal unintentionally sends information about its future movements from its physical structure and past activities. One might say it has certain personality trait properties about its potential and actual motions. Some of these include motions that attempt to minimize predictability, while still being effectively honest in their communicative nature. However, each agent also has movements, whose evolutionary function is to increase or decrease distance through purposive, though not consciously intentional, miscommunications—to lead the other agent in the wrong direction, which will also have the net effect of increasing or decreasing distance. These latter behaviors are purposively communicative, because they are directed at the other agent, but they are a deceptive sort of communication whose function is to misdirect the other agent from the actual common world that the two agents share to an illusory common world. However, these attempts at deceit in communication could not occur if it were not for the fact that other more accurate information about the animal's activities into the future were not also being transmitted. Without information that can be interpreted as a "true signal" of a common world, the miscommunication or "false" *signal* could not occur. Moreover, even "false" signals are only false if interpreted in a certain way based on their apparent meaning, just as "true" signals are only true if the perceiver of the signal

understands the standard meaning of the signal. But once the receiving animal acquires knowledge of the tendency to use deceptive communications on the part of the sender, then this misinformation can be interpreted correctly, thus actually providing true information of the sender's intentions (cf. Oberzaucher and Grammer this volume).

Au: pls check.

Common worlds of a competitive sort occur more often within species than between species. And within species the dialogical nature of the competitive dance is even more highly structured and communicative than between species. Whereas, in the case of between species competition of the sort just described information interpreted as a "true signal" does not have an original communicative purpose while the false signal does, it is often the case within species that both the *true and false signals* have been originally selected as communicative in purpose.

Territoriality is an important form of adaptation within many species, whether walruses, wolves, frogs, birds, or humans. In all of these cases and many more, signals are used in a positive communicative sense to indicate territorial boundaries (Wilson 1975). These signals, which have conventional meanings understood by participants, are essentially communal and globally communicative—they signal to each other where each agent is and intends to stay. The signals are essentially honest. For instance, large, older bullfrogs are louder and deeper in their croaking than young frogs, so that other male frogs give them wider berths, while females may approach rather than avoid sounds of deep croaking (Gerhardt and Huber 2002). Since the signal's quality typically depends on the size and age of the frog, it is difficult to fake (Maynard Smith 1984). The signal here is tied directly to physical parameters of the frog that are relevant to territoriality and the likelihood that the sender can defend the territory. When territorial signals are honest like this, there is little reason for competition to lead to combat, where rivals typically have more to lose than to gain by direct contact. By creating a common world of communicative signals that each can send and receive with apparent understanding, they limit the need for dangerous forms of competitive activity. Communication replaces contact. But, I think it would be useful to look a little more closely into this process of communication between within-species competitors of this sort and the common worlds that are involved.

Whereas, in our example of predatory-prey relations there is nothing to be gained from honest communication of a purposive sort, this is not the case in within-species territorial communications. This is not to say that honest communications cannot exist between species. In symbiotic relationships between species, individuals of different species cooperate with each other to pursue their own ends (Wilson 1975). Cleaner fish and their larger host fish communicate honestly to each other in working out when it is safe for the cleaner fish to approach and clean the larger fish (Trivers 1971; Grutter 2004). However, such cooperative ventures with purposively honest communications between individuals are atypical between species. Honest communications are much more likely to occur within species as adaptations that collectively increase reproductive success of members of the species, relative to members of other species with fewer cooperative strategies. However, such success does not depend on group selection but on the emergence of evolutionary stable strategies (Maynard Smith 1982). These behavioral strategies that are widely adopted and stabilized within the species result in types of common

worlds where honest communications play an important role for ends that can be described as both collective and individual, but where deceit is also a subsidiary strategy associated with individual selective processes (e.g. Trivers 1971; Maynard Smith 1984). Thus in the case of territoriality, the use of common signals across all members of the species to indicate where an individual or a group resides, reduces the costs of competition between members of the species by comparison with other species without those signals. It does not prevent competition, itself, from occurring, or prevent a certain amount of deceptive signaling from occurring. However, if the cost of deception were as cheap as the cost of honest signaling, then the honest signal would disappear (Maynard Smith 1984). So, honest signaling has to be solidly grounded in intrinsic properties associated with the signal, or in close relationships among individuals whose genes gain advantage from the signal. In the case of bullfrogs, it is the size and age of the frog that is the honest basis of the signal and is relatively difficult to fake.

Communicative signals of location can serve multiple functions in the common worlds of a species. Not only do they indicate the location of individuals and groups for the purposes of territorial competition, but they can also identify individuals and groups as “friends” or “foes”. Among closely tied individuals they can indicate where particular individuals are, who desire to join together. The extent of such identification of friends and foes by signals can be extensive. For instance, in the case of whales (Rendell and Whitehead 2005) it has been shown that communalities in clicking signals, which can be heard across vast distances in the ocean, are used to identify individuals and groups that share large feeding areas. Each large language group can consist of thousands of individual whales that share regions on an annual basis. How such huge collective dialects could have evolved is still under investigation, but it is an indication of how signals can evolve into a basis for cooperation among vast numbers of individuals, who do not know each other as individuals but who nevertheless cooperate with each other through signals based on a common linguistic heritage. The analogy of whale communication to how differences in human languages and dialects can serve as a means of identifying friend and foe is not, I think, illusory.

2.5 Cooperation and communication in communal groups

Thus far we have considered personal and common worlds between individuals and groups mainly in competitive situations between different species in predator–prey relations, and with respect to the role that communication serves in reducing confrontations with high energetic costs among individuals within a species. In these cases the amount of communality of personal and common worlds tends to be limited. This is especially the case in interspecies relations.

Cooperation can change all this. Especially when cooperation occurs within a species there is a vast area of similarity in personal worlds, which can provide the potential for highly articulated common worlds among cooperating individuals. Because the sensory organs, bodily structure, brains, and behavioral capacities of individuals of the same species are built on a single general model, the personal worlds of adult members of the species, who live together, can have a high degree of overlap in intentional relations. As a

result of contingent overlap of intentional relations, they can share common worlds, even when they are not acting together toward common goals. If there is a sudden noise nearby, most if not all of the individuals in a group will hear it and orient toward it. Depending on the noise they will tend to act in the same way either approaching or escaping from its apparent source, which will be understood in the same way by nearly all the individuals. In the case of perceiving a food resource if they are in a state that can use the energy from the resource, they will all be attracted to the resource and consume it. In general each individual's spontaneous behavior in response to the situation will tend to be similar to the behavior of their peers in shared situations. However, an individual's behavior can also be primed by the behavior of their peers, which can produce interlocking personal worlds and more extensive communication-based common worlds. For instance, consider the coupled behavior of schools of fish, or in the flight of birds in flocks. There is a sense in which these cases can be viewed as cooperative, but they need not be intentionally cooperative. In order to avoid predators, individual fish in the school may direct their behavior primarily at being toward the middle of the school, but in the process they must also respond synchronically to the behavior of other fish around them, otherwise they would crash into each other and the gain from schooling would be lost (Hamilton 1971; Dawkins 1976). Conformity as a result of priming like this may be a general basis for action tendencies involving close encounters with others that applies throughout the animal kingdom, and not just in schools of fish.

Cooperation in examples like schooling seems to emerge out of individual action in shared situations. Even more complicated forms of apparent cooperation and common worlds, but not involving priming of coupled similar behavior, may also take this form. Consider the example of cooperative hunting in chimpanzees. While there is some evidence for purposive cooperation, for instance on occasion there seems to be some division of labor in achieving the goal, for the most part such hunting can be interpreted as an individualistic affair (Cheney and Seyfarth 1990; Tomasello *et al.* 2005a; but see Boesch 2005 for an intentionally cooperative account of chimp hunting, and Tomasello *et al.* 2005b for a reply). For example, in attempting to capture a monkey, each chimp will approach the monkey from its own location and try to capture it. In pursuit of the monkey they will eventually surround it. However, it is probably not the case that the chimps purposely surround the monkey in order to capture it collectively, but rather that each pursues the monkey opportunistically from their currently position, while maintaining some distance from other chimps so as to have a clear shot at the monkey. As the chimps scramble around in pursuit of the monkey, eventually they achieve a collective position, where the monkey cannot escape. Thus the apparent collective or joint action emerges out of individual pursuits, but with concurrent mutual awareness of each other and adaptation of actions in response to each other. This is very different from planned action of the same sort, where humans may try to corner an animal by directing each other through intentional communications both verbal and non-verbal to adopt specific positions relative to each other, or to set a trap and chase the animal into it. The kinds of common worlds that emerge from similarity in personal worlds of the chimps in common situations, are not *intentionally constructed*, and do not involve common goals mutually

understood, as do cases of human joint action that have similar outcomes. In the human case the common world is understood by each individual as shared with others, but this is probably not so in the case of the chimps. This distinction is crucial and relates to the notion of “common ground” used by philosophers and psychologists of language (Clark 1996).

Before turning to these more complex forms of cooperation in humans and into a discussion of common ground that depends on human forms of communication, we need to consider the role that intentional communication plays in the formation of cooperative common worlds in simpler species. It will be recalled that we defined communication of a minimal sort as a causal relation between the agents, where one agent affects the behavior of the other. Using the term in this minimal way, even the behavior of schools of fish, or of chimps chasing a monkey, involve communication. But what I wish to consider here are more complex forms of learning from others that involve the use of signals in a cooperative manner in communal groups. These signals have meanings that are understood by all; hence, like the territorial signals that we have previously discussed, they purposely communicate information about common worlds for actions and interactions.

A number of species have communal signals that reliably differentiate situations, some involving dangers, for instance alarm calls, and some regulating cooperative social interactions such as approach grunts, begging gestures, and play faces (Cheney and Seyfarth 1990; Seyfarth and Cheney 2003). In some cases, like vervet monkeys, there can be specific alarm calls for particular dangers that are standard across the species. Although they have these calls in their repertoires, young monkeys must learn the situations that are appropriate for each of the calls. One result of this learning aspect is that young monkeys may use the wrong calls for particular dangers, which, if accepted by others, would lead to inappropriate defensive behavior. However, adult monkeys adapt to variations in accuracy of the calls from particular source individuals by not responding immediately to suspect calls and investigating the situation instead. Only if they or more reliable sources detect the danger will they respond with further calls and defensive behavior (Cheney and Seyfarth 1990).

In cases like alarm calls it is not clear to what extent they can be viewed as intentionally communicative. They often operate more as expressive communications of a non-purposive sort, for instance screeching with fear or pain. However, expressive emotions, even in these cases, are likely to have communicative purposes of a general nature, since they tend to occur more often in contexts where other individuals in one's group can take notice of them and less so or not at all when alone. And some, like expressions of anger, are directed at particular individuals within groups. In cases like vervet monkeys, where some signals have specific meaning, the specificity of the signal seems to imply purposive communication of a meaning to others, even though it does not involve consciously intentional communication (Seyfarth and Cheney 2003). Again one way to tell whether it is purposively communicative is if having an audience or not makes a difference in expression. There is an energetic cost to signaling, as well as a danger to signaling that a predator can use to locate the signaler. Hence, signaling is unlikely to occur in the absence of social contexts. However, because monkeys tend to live in groups, an audience is almost always there, so for some signals selection may not have differentiated these

two conditions. Nevertheless, sometimes monkeys can be very sensitive to the presence of an audience. Recent research involving langur monkeys has found that a dominant male monkey will continue to provide a predator alarm call until it receives an echoing response from *all* members in his group, guaranteeing that they have all heard his message (Zuberbühler 2006). The functional purpose, and perhaps even the monkey's goal in the continuing signal here seems to be to insure that all group members, composed mainly of relatives, have made a response to the danger, and it is willing to put itself in danger in order to insure a response to the danger from all members in its group.

While alarm signals and some emotional expressions are not usually intended to communicate to particular others, play faces, begging gestures, and approach grunts certainly are directed at particular individuals and may be considered intentionally communicative. The boundary between purposive and intentional communication begins to get fuzzy in situations like these. In begging encounters, the agent begging is engaging in the behavior in the hope of obtaining food, even if the performance is species-specific, and adapted to this purpose. Both the agent begging and the other agent might be said to know the meaning of the gesture, although the form of their knowledge is not the same (Barresi and Moore 1996). For the individual producing the gesture, its use is directly controlled by the desire for food from the other, with minimal understanding of how the gesture works to achieve this goal, while for the other agent the perception of the gesture indicates further behavior that can be expected from the begging agent in its attempt to acquire some of the food.

Even deceptive gestures between species can sometimes operate this way. Consider the false broken wing display that some birds use to attract predators away from nests of eggs (Ristau 1991). The agent here seems to have the immediate goal of directing the predator away from the nest, and uses the species-specific behavior to achieve this goal. But the bird is not consciously aware that the wing display falsely communicates the status of its ability to escape the predator in the end. The bird can only use this wing display to miscommunicate in limited contexts, so it is not a signal it can use intentionally to achieve different purposes. Still, the point here is that despite these limitations on intentional communication or in false communication, organisms can engage in intentionally communicative actions to achieve particular goals, though they are not consciously aware of how their communicative gestures operate to achieve these goals.

In some cases the communicative gestures that are used are not species-specific in meaning, but are learned arbitrary signals that operate in specific relationships with others. Learning such arbitrary signals may seem analogous to learning human language, but the form of learning is different. Apparently, chimps develop such idiosyncratic signals through an individualistic learning process (Tomasello *et al.* 1994), though there is also evidence of group-specific or cultural use of gestures, which may have their source in mimetic processes more similar to human language and possibly transitional to it (Pollick and de Waal 2007). In the individualistic process, a behavior may accidentally precede a desired outcome, for instance obtaining some food from another individual; then the behavior is produced again in a comparable situation, when that outcome is desired. If it works again, the behavior is reinforced, and produced once again.

Both agents come to understand that the gesture plays a communicative role here to achieve the purpose, though their understanding of this role may differ. While the sender may perceive its success in using the signaling behavior to achieve its goal and the receiver may use the signal to indicate the sender's desires, they need not share a common interpretation of the signal. Deceptive signals can be reinforced in this way as well. An accidentally generated alarm call that sends others away from a food resource may develop into a goal-directed behavior to use when one wants a food resource and others are preventing one from getting it. But in such a case it is the outcome of the behavior that reinforces it in the sender, not the understanding that a deceptive communication is involved (see Barresi and Moore, 1996, Tomasello 1998, and Seyfarth and Cheney 2003, for further discussion of differences in interpretation of behavior between animal actors and observers).

In the examples that we have considered so far, the communicative signal functionally creates a common world (the meaning of the signal) for several agents in the absence of a conscious understanding of the meaning of the signal. In the case of false communication, it is an illusory common world that is communicated, one with the function of generating in the receiving agent a misrepresentation of the actual common world. While in some cases these signals seem purposively communicative, in none of them is there a hint that the signal's meaning is understood *as such* or that it is *consciously* used to communicate information about some aspect of the external world in order to achieve a joint goal.

In Barresi and Moore (1996), Chris Moore and I discuss one example that comes close to a purposive communication about a perceived common world between individuals in order to achieve a joint or common goal, where the meaning of the signal may be understood. As originally described the case involved a vervet call that precedes entry into an open area that might potentially have a predator (Cheney and Seyfarth 1990; Dennett 1987). Subsequent, more detailed research has been conducted on a related "move grunt" in baboons (Rendell *et al.* 1999; Seyfarth and Cheney 2003). In this communication, monkeys (or baboons) in the group use the call apparently to indicate to each other both that they are ready to enter an open area and that it seems safe to enter the area. It is as if one monkey says, "I'm ready to go and don't see any danger, do you?" and another responds, "I'm ready to go too and don't see any danger either." Following multiple communications of this type, the group of monkeys enters the open field.

In this example, it seems that each monkey directs its signal to the other monkeys and also keeps track of where they are in the communication. If a sufficient number of the monkeys agree that the timing is right and the situation is a safe one, they move forward, but if only one or a few monkeys give the signal they do not enter the open area. What makes this case interesting is that there is a common goal governing the communication and that the communication is about a potential joint action as well as an objective situation external to the monkeys' immediate relations to each other. The several agents seem knowingly to work together to determine whether they are ready to leave and whether it is safe to enter the open area, and they communicate to each other their assessment of time and situation. Only if most of them agree do they move forward. If this interpretation of what is happening here is correct, it seems natural to conceive of this situation as

one of joint action and intentional communication about a common world that they share in support of that action, something that seems to occur quite rarely outside of the human species, but quite often within it.

2.6 Common ground, common goals and common worlds: language and communication from bees to humans

In his book: *Using language*, Herb Clark (1996) gives great importance to the notion of *common ground* in linguistic and non-linguistic communication. The basic idea is that typically when individuals communicate with each other it is in a context where each individual knows, or at least presupposes, a lot about what the other individual knows prior to the context of their present conversation, as well as what is being shared with them immediately in the present context. *Common ground* is the content of this presupposed information that contextualizes the conversation. It is the shared or common world in which the conversation occurs—a world that is presupposed by the interacting agents in their communications. The agents then use their commonly understood language to communicate about this common world in order to modify it or elaborate on it in some way.

Leaving aside issues associated with the high level of reflexive self-knowledge that might appear in human examples of common ground, we can wonder about the extent to which something akin to common ground is found in animal communication. The example of vervet monkeys (or baboons) entering an open area would seem to involve something like common ground since it is a context of the communication that each monkey is in some way aware that the other monkey shares with it a common goal of possibly entering the open area, and sees their communication as a means of working out whether they should. However, attributing awareness in a reflexive sense to this common knowledge is probably going too far. The communicative signal here is species-specific for the task at hand, and may be elicited directly by the complex of intentional relations involving several monkeys aware of each other as they approach an open area. Still, it is difficult not to see a strong analogy here to common ground. There is a common contextual world that elicits the communication and includes some awareness in the monkeys of the function of the communicative exchange in jointly assessing properties of the common world that the monkeys share. Even if it is all wired up to operate in specific contexts, it still succeeds in doing so because the personal worlds of the individuals overlap and intertwine into an appropriate common world in the context.

The famous waggle dance of honeybees provides another illuminating example of something akin to common ground in the animal kingdom, but on a much lower order with respect to cognitive mechanisms (von Frisch 1953; Wilson 1975). Honeybees are a species with a high level of common worlds relative to individual personal worlds because their highly cooperative communal life is strongly selected for through kin selection (Wilson 1975; Trivers and Hare, 1976). Personal reproductive success plays second fiddle to inclusive fitness of the kin group involved in the common hive. So each individual acts in functional roles with respect to the cooperative reproductive venture of the whole group. The waggle dance is a case in point. What it achieves is the communication

of locations that other members of the hive can observe and use to decide where to search for food to bring back to the hive to feed the queen's offspring. While I would not want to suggest that the bees have any idea of what they are about in terms of intentional communication through a common language or awareness of common goals, there is little doubt that there exists a highly articulated common world of shared meaning of the innately governed waggle dance, and common goals associated with the hive, which dominates behavior associated with the personal worlds of individual bees. As such it provides an impressive model analogous, though on a very limited scale, to human society as well as to what we might hope to achieve in social robotics. Each organism has its own goals, but they are coordinated with the goals of others, so that innately determined common goals are achieved. Individuals also communicate through the intricately constructed waggle dance just the right information to their kin, to help maximize the discovery of and utilization of energetic resources that can increase reproductive success of the hive, relative to a species without such a useful communicative strategy.

If we compare humans to these other animals several important differences appear. First, whereas the communicative signals of monkeys and bees, as well as of most other animals, tend to be fixed and innate, most human signals are arbitrary and learned. Grice (1957) made the distinction between natural sign and non-natural sign or signal, where a natural sign is something like expressive behavior that could indicate a fixed meaning like a particular emotion, while a non-natural sign would be language or a gesture like pointing that could have arbitrary meaning. However, animal communication depicts a middle kind of communicative signal. The waggle dance is arbitrary, in that the relationship between what is expressed and the signal has no necessary relation in the way that emotional expression, or physical size of an organism, might directly express its meaning. However, on the other hand, the dance is hard-wired, and not learned. Indeed, if the bee has any awareness of it at all, it would probably feel much like what we would consider expressive emotional behavior in humans. So it has this kind of hard-wired property of natural signs, which necessarily have the meaning that they have, rather than being entirely arbitrary like human language, where words can have different meanings in different languages and over time in the same language.

A second important difference is that while human communication using verbal language and other communicative gestures, can have a wide variety of shared meanings generated from multiple signs and their combinations, the examples of animal communication that we have considered here, are restricted to a very limited vocabulary and content of the communications (Wilson 1975; but see Marler 1965, on the complex mixture of signals of different types in social communication in primates). The common worlds that appear in the content of animal communications are highly restricted compared to the common ground and common worlds of human communication. And related to this is the fact that joint action on the part of animals tends to be limited to innately determined coordinated action, where each individual tends to pursue its own personal goals, which complement, and depend upon the personal goals of other individuals. Communication involving signals of a complex sort like the waggle dance plays only a small role in the coordinated activity in innately communal species like bees. By contrast,

humans constantly create, through the generative power of language, new forms of cooperative joint activities and forms of social life that can extend over vast numbers of human beings and over many generations.

However, the power of human communicative systems is also associated with a cost. In contrast to the cooperation that occurs among bees, where individuals automatically focus on group goals in pursuit of individual interests and in their communications to each other, human cooperation is much more fragile. Individual self-interest and personal worlds that do not coordinate with other individuals to form common goals and common worlds, is much more likely in humans. So joining with others in common pursuits requires both learning about the interests of others, and developing common interests with them. And the generation of joint activities with common goals and worlds for self and other, is a constant challenge played off against the pursuit of individual goals that often involves taking competitive advantage of the other.

Thus, although the human world has some strong analogies with cooperative activity in bees, it has much more in common with the mix of competitive and cooperative activity in wolves and whales, monkeys, and chimps. Dishonest communication, providing misrepresentations of intentions to pursue common interests, is a daily occurrence in human lives. Furthermore, because of human self-reflexive capacities, these deceptive activities are consciously intended to deceive. And humans have even gained control over what was probably an originally honest emotionally expressive system, in order to lie with smiles, and encourage with frowns. So natural signs like emotional expressions have coevolved with verbal communication to provide concurrent information about intentions of others, which can lie as well as tell the truth. However, probably because of their original connection to natural signs, emotional expressions are more difficult to manipulate than verbal language in dishonest communication, and thus serve as a more useful means to detect honest communication than verbal language (Ekman 1972; Ekman *et al.* 2004; Parkinson 2005 Oberzaucher and Grammer this Volume).

2.7 The role of communication in the construction of common worlds in human development

In our early discussion of personal and common worlds we noted that common worlds are typically constructed out of overlap and interconnection of personal worlds of embodied agents as they relate to each other and to the surrounding natural world. This definition is particularly useful for describing cases of competitive relations between organisms of different species. However, with increasing dependence on social life that involves learning from and communicating with others, personal worlds become so intertwined and co-constructed with common worlds that personal and common worlds relate less to the natural world than to one's shared social world. In humans, socially constructed common worlds that include a materially reconstructed natural world, largely determine the construction of personal worlds.

From early infancy onward, human beings are motivated to be social and to share common worlds with other humans (see Barresi and Moore 1996 and, especially,

Tomasello, *et al.* 2005a, 2005b for reviews of human social-cognitive development and how it differs from other organisms; also Sebanz and Knoblich this volume, and Prinz this volume). Both adults and babies work at creating interpersonal worlds. For its part the baby shows a strong interest in the human face and has a tendency to smile in response to adult smiles. Signs of mimicry of observed facial expressions occur immediately after birth and occur frequently by the third month when interpersonal mimicry between adult and baby reaches an almost game status, where they take turns copying each other, and where the baby may even test the adult's ability to copy it (Nadel 2002; Striano *et al.* 2005). On the adult's part, there is a continuing attempt to engage the baby through various means of drawing its attention to the adult and to its own body parts, and to objects that the baby can investigate jointly with the adult (Adamson and Bakeman 1991; Moran *et al.* 1987). These complementary activities between the baby and adult lead to the construction of common worlds, usually of short duration of shared attention to each other and to the immediate environment. During these periods of mutual attention, expressive communications of gestures involving face and body parts as well as sounds, are exchanged and often repeated between the adult and baby through mimicry. Toward the end of the first year, imitation comes to play a role where actions toward particular objects are copied between the adult and infant, and attention is directed by one or the other to particular objects, through verbal emphasis, pointing gestures, and shifts in visual direction (Adamson and Bakeman, 1991; Bates 1979; Trevarthen and Hubley 1978). The concurrence of these shared intentional relations tend to occur in activities involving play with objects, where these joint activities that now involve objects build on the already existing ability to pay mutual attention to each other's actions and to mimic or imitate those actions.

Common ground plays an important role in the development of joint activities. The infant becomes reflexively aware of being in a common state of attention with the adult or not, and can draw the adult's attention to some activity or object when the adult is not attending. Or the baby can shift its own attention to the direction of orientation of the adult, or to the direction of a pointing gesture made by the adult. Although the infant is not yet reflexively aware of itself as distinct from the unit of adult and self, it is aware of their congruence in attention to some referential activity (Gomez 1994, 2004; Moore and Barresi, in press).

However, during the second year of life the baby gradually develops the ability to distinguish self from other, and to become aware of differences in the intentional relations of self and other (Barresi and Moore 1996; Tomasello *et al.* 2005a; Moore and Barresi, in press). It can join in with the activity of the other, not as it did earlier on, as a joint activity, but more imaginatively (Hoffman 1977; Tomasello *et al.* 2005a). It also becomes aware of how the other individual is responding to its own activity, now, not in terms of joint goals, but as another individual's perspective on self (Lewis *et al.* 1989). These developments involving distinguishing self and other in separate intentional relations seem to co-occur at the same time that the infant becomes clearly aware of itself as distinct from other, as indicated in the mirror self-recognition task (Amsterdam 1972). So the infant becomes able not only to share common worlds involving joint object

directed activities with others, but can now reflexively distinguish its own personal world of object directed activities from those of others, and intentionally co-construct common worlds through communication of intentions. Thus the infant has awareness of the common grounds that are shared with others and can intentionally manipulate them through communicative actions.

In the process of coming to distinguish personal and common worlds, and in using communication as a means of moving from one to the other, the infant learns skills in mind reading as well as skills in communicating. Verbal language conveniently comes to the fore at this time, to make communication of possible joint activities, and of personal and common worlds, much more efficient than through expressive communications and gestures (Bates 1990; Moore 2006; Tomasello 2003). Thus the number of possible common worlds expands enormously through referential language, which can be used to generate all sorts of joint activity and common understandings between the child and others. Some of these common worlds are simulated in play activity, while others are directly acted through imitated and coordinated activities.

Through such processes where verbal and non-verbal communication plays a constant ongoing role, and facilitated through formal education where communication is systematically used to transmit cultural knowledge, the child eventually acquires the common worlds of family and culture and can find its own roles in these shared world activities. Eventually, as an adult, while pursuing his or her own personal goals the individual can also share in the co-construction of new common worlds in relationships with particular others and in more general cultural activities. The end result of this process is an ongoing human society that progresses culturally through the creative efforts of individuals and groups who continuously construct new personal and common worlds. Without the ability to understand the personal worlds of self and others and to work together to continuously reconstruct common worlds, human progress would be impossible. With it, human society far outstrips other animal societies in being able to modify, and hopefully improve, the form of their adaptation to physical and social reality.

2.8 Prospects for the Cyberiad, Mark II

About 20 years ago, I proposed what I called the “Cyberiad test” as a means for determining potential limits on human self-knowledge (Barresi 1987). The basic idea for the test came from Stanislaw Lem’s novel, *The Cyberiad* (1976), which depicts a future time when humans no longer exist, but robots come to spread throughout the galaxy. These robots had their origin in human construction, but are now engaged in self-construction. There are particular engineer robots that construct robots ordered by other robots, and in this manner create new types, and ever more robots, some of which adapt and persist, while others do not. The Cyberiad test that I proposed involved imagining a future when we would know enough about ourselves that we could consciously initiate something like a Cyberiad. The notion was to replace a society of humans with a society of their social equivalents that are cybernetic robots. If we knew enough about ourselves, our biological, psychological, and social nature, we could construct functional copies of the individual humans in a society and create an equivalent society of robots that could carry on as the

humans would. One of the main activities of the society would be to construct new robots to replace previous ones through a process analogous to human reproduction. The test was to see if we could imagine having enough knowledge of human nature that we could be confident that we could create such a society so strongly analogous to human society that it could survive as long as human society might be expected to survive as a species, say a million years.

The Cyberiad test was constructed as an alternative to the Turing test (Turing 1950), which tests the capacity of a program to simulate human intelligence, but is limited to a test of the program's ability to communicate through language to humans and not be distinguishable in linguistic responses to those given by actual humans. One of the issues that bothered me at the time was the then current research focus on representational approaches to human intelligence and its general application to computer intelligence as it was ensconced in the Turing test. As part of the article, I critiqued this notion of an essentially disembodied mind associated with the Turing test, whose contact with and grounding in the physical world was left in doubt. I contrasted the Turing test with the Cyberiad test, which was an embodied intelligence test in the natural and social world, for which more than disembodied programs capable only of language would be required to indicate natural intelligence.

Since the publication of my article, though not because of it, artificial intelligence has turned away from its reliance on representational approaches and has come to focus more on robots learning directly about the physical and social world from acting on it or interacting with it and acquiring knowledge out of these interactions (Brooks 1991; Clark 1997). This shifts the emphasis from "accurate" representations, to sensory-motor loops of increasing power and flexibility, to quickly assess properties of the environment, and to act in them. And instead of complex representational networks that model the world as a whole, it shifts knowledge to one of learning adaptive responses to particular situations that can be recognized directly based on sensing and categorizing cues, and only rarely using elaborate representations. A similar shift has also occurred in human psychology, which has turned away from representationalism and methodological solipsism (Fodor 1980) to focus more on a more Gibsonian (Gibson 1979) naturalistic psychology of the informational basis of perception and action, both in non-social and social activity. The embodied approach to communication and action in humans and robots represented in this volume is an ideal example of this shift away from representationalism.

By bringing robots and humans as embodied agents into a single physical world, it is possible to see analogies that might bridge quite different kinds of embodied agents, without having to bridge them as one kind of representational mind to another kind of representational mind. Whatever the differences that there might be between robot and human minds, or the minds of humans and other organisms, as embodied agents we all exist in the same physical world, and can exchange information between each other through one or another form of communication involving physical signals as I have tried to describe it in the present chapter. The question I would like to propose now is how this communication and action approach, which actually harks back to the origins of cybernetics (Wiener 1948), can be used to pass a weakened version of the Cyberiad test.

Instead of aiming directly for the gold standard of a Cyberiad that matches human intelligence and social life, perhaps we should consider the possibility of aiming initially for the knowledge necessary to create a bee-like Cyberiad. How would we go about creating a society of robots that can reproduce themselves for multiple generations on the analogy of bee society? One big advantage of using bees as a model is that we would not have to deal with cultural and historical change in social structure, or with human creativity in generating new forms of knowledge and common worlds, two problems that I think may be difficult for us to solve at our present stage of knowledge of human nature. It would also avoid issues associated with competition among individuals within the social group or with complex forms of social learning. If we were to succeed in generating a model of a Bee-Cyberiad, we could then approach these more difficult issues, by focusing on the social structures of more complex social species of animals, like wolves and whales, to see if we could integrate the more competitive and flexible adaptations of such species into a basically cooperative model of reproduction created on the bee model. Eventually, after considering monkeys as well as other higher primates we could return on a holistic level to the problem of a human-like Cyberiad. But, even here, we could articulate the problem in such a way that it isn't a matter of recreating human society in robotic form, but of matching human-like society in a different form of embodied agent (cf. Stephan *et al.* this volume).

My main goal in suggesting such a revised research proposal is that we should try to develop a language of embodied agents and of their abilities to engage in social life that cuts across types of agents, types of personal and common worlds, types of information processing, communication, and common ground, to root principles that are both physical and can be interpreted at a social level. In other words, what we need is a kind of universal language of social life, one perhaps not very different from the one I have tried to develop in this chapter. With such a language, perhaps we can develop enough knowledge of the laws of social nature that we can come close to understanding our own embodied existence as human beings, and, indeed, pass something close to a full-strength Cyberiad test.

2.9 Conclusion

In this chapter I have attempted to investigate some boundary conditions on embodied communication and its role in social life. The types of embodiment and communication that I have considered have been very general—general enough to include a wide range of animal species and robots among embodied agents, and causal relations that can be interpreted as involving information and as communicative between agents at far distances in space and time. However, they have not been perfectly general. I have had nothing of a systematic nature to say about possible communication between embodied agents of species from different planets, whose forms of embodiment, psychology, biology, and social nature might be radically different from each other. I've also said little about quasi-embodied agents, or of relations between humans and robotic agents, whether fully embodied or not. I understand that much of what is going on now in robotics depends

on teaching robots how to copy humans or interact with humans and, through this process, develop their own skills at human-like behaviors. I take this as an exciting area of research but one which I have not explored (see Kopp *et al.* this volume; Galuntacci and Steels this volume; Stephan *et al.* this volume).

What I have done is to look at some basic principles that apply to embodied individuals and the forms of communication and types of social arrangements that exist between them both within and between species here on Earth. I've also tried to provide a language for discussing these basic principles and variations in their application across species. My hope is that in developing a common language and in searching for some basic principles of embodied social life, we can come to a better understanding of our own embodiment as humans, and of our own form of social life. I've also briefly considered some of the implications of these principles and forms of social life for the development of the field of social robotics. Much more needs to be done in the search for a common language of embodied social life and in determining the basic principles of embodied communication involved in social life. I can only hope that the present chapter provides at least a small initial attempt at developing such a common framework.

Acknowledgements

The present chapter was written in the context of the multiple presentations that I witnessed during the ZiF year on Embodied Communication and I would like to thank all of the participants, and in particular the organizers, for their contributions to my thinking on the topic of embodied communication in humans and machines. Although I do not reference the particular talks, I hope that some of these contexts will become apparent to anyone reading other chapters in this book. In addition to my gratitude at being invited to participate as a fellow in the ZiF year, I wish to thank the Social Science and Humanities Council of Canada, and Dalhousie University, for support that made it possible for me to participate in the ZiF year and to work on the present project. Finally, I would like to give special thanks to Chris Frith, Scott Jordan, and Guenther Knoblich for their very useful recommendations for revisions from an earlier version of this chapter.

References

- Adamson L and Bakeman R (1991). The development of shared attention in infancy. In R Vasta, ed. *Annals of Child Development*, Vol. 8, pp. 1–41. London, Jessica Kingsley Publishers.
- Amsterdam BK (1972). Mirror self-image reactions before age two. *Developmental Psychology*, 5, 297–305.
- Barresi J (1987). Prospects for the Cyberiad: certain limits on human self-knowledge in this cybernetic age. *Journal for the Theory of Social Behaviour*, 17, 19–46.
- Barresi J (1999). On becoming a person. *Philosophical Psychology*, 12, 79–98.
- Barresi J (2007). Consciousness and intentionality. *Journal of Consciousness Studies*, 14, 77–93.
- Barresi J and Moore C (1996). Intentional relations and social understanding. *Behavioral and Brain Sciences*, 19, 107–22.
- Barresi J and Moore C (2008). The neuroscience of social understanding. In J Zlatev, T Racine, C Sinha and E Itkonen, eds. *The Shared Mind: perspectives on intersubjectivity*. 1, pp. 39–66 Amsterdam/Philadelphia, John Benjamins.

- Bates E (1979). Intentions, conventions, and symbols. In E Bates, L Benigni, I Bretherton, L Camaioni and V Volterra, eds. *The Emergence of Symbols: cognition and communication in infancy*. New York, Academic Press.
- Bates E (1990). Language about you and me: pronominal reference and the emerging concept of self. In D Cicchetti and M Beeghly, eds. *The Self in Transition: infancy to childhood*, pp. 165–82. Chicago, Chicago University Press.
- Boesch C (2005). Joint cooperative hunting among wild chimpanzees: taking natural observations seriously. *Behavioral and Brain Sciences*, **28**, 692–3.
- Brentano F (1874/1973). *Psychology from an Empirical Standpoint* London, Routledge and Kegan Paul.
- Brooks R (1991). Intelligence without representation. *Artificial Intelligence*, **47**, 139–59.
- Cheney DL and Seyforth RM (1990). *How Monkeys See the World: inside the mind of another species*. Chicago, University of Chicago Press.
- Clark A (1997). *Being There*. Cambridge, MA, MIT Press.
- Clark HH (1996). *Using Language*. Cambridge, Cambridge University Press.
- Dawkins R (1976). *The Selfish Gene*. Oxford, Oxford University Press.
- Dennett DC (1987). *The Intentional Stance*. Cambridge, MA, MIT Press.
- Ekman P (1972). Universals and cultural differences in facial expressions of emotion. In J Cole, ed. *Nebraska Symposium on Motivation*, pp. 207–83. Lincoln, University of Nebraska Press.
- Ekman P, Campos JJ, Davidson RJ, and de Waal FBM, eds. (2004). *Emotions Inside Out: 130 years after Darwin's the expression of the emotions in man and animals*. New York, New York Academy of Science.
- Fodor J (1980). Methodological solipsism considered as a research strategy in cognitive psychology. *Behavioral and Brain Sciences*, **3**, 63–110.
- Gerhardt HC and Huber F (2002). *Acoustic Communication in Insects and Anurans: Common Problems and Diverse Solutions*. Chicago, University Chicago Press.
- Gibson JJ (1979). *The Ecological Approach to Visual Perception*. Boston, Houghton Mifflin.
- Gomez JC (1994). Mutual awareness in primate communication: a Gricean approach. In ST Parker, RW Mitchell, and ML Boccia, eds. *Self-Recognition and Awareness in Apes, Monkeys, and Children*, pp. 547–73. Cambridge, Cambridge University Press.
- Gomez JC (2004). *Apes, Monkeys, Children, and Growth of Mind*. Cambridge, Harvard University Press.
- Grice HP (1957). Meaning. *Philosophical Review*, **66**, 377–88.
- Grutter A (2004). Cleaner fish use tactile dancing behavior as a preconflict management strategy. *Current Biology*, **14**, 1080–3.
- Hamilton WD (1971). Geometry for the selfish herd. *Journal of Theoretical Biology*, **31**, 295–311.
- Hoffman M (1977). Empathy: its development and prosocial implications. In CB Keasey, ed. *Nebraska Symposium on Motivation, Vol. 25: social cognitive development*, pp. 169–217. Lincoln, University of Nebraska Press.
- Lem S (1976). *The Cyberiad: Fables for the Cybernetic Age*. New York, Avon Books.
- Lewis M, Sullivan MW, Stranger C and Weiss M (1989). Self-development and self-conscious emotions. *Child Development*, **60**, 146–56.
- Marler PR (1965). Communication in monkeys and apes. In I DeVore, ed. *Primate Behavior: field studies of monkeys and apes*, pp. 544–84. New York, Holt, Rinehart and Winston.
- Maynard Smith J (1982). *Game Theory and the Evolution of Games*. Cambridge, Cambridge University Press.
- Maynard Smith J (1984). Evolution and the theory of behaviour. *Behavioral and Brain Sciences*, **7**, 95–125.
- Moore C (2006). *The Development of Commonsense Psychology*. Mahwah, NJ, Lawrence Erlbaum Associates, Publishers.

- Moore C and Barresi J (in press). The construction of commonsense psychology in infancy. In P Zelazo, M Chandler and E Crone, eds. *Developmental Social Cognitive Neuroscience*. New York: Psychology Press.
- Moran G, Krupka A, Tutton A and Symons D (1987). Patterns of maternal and infant imitation during play. *Infant Behavior and Development*, **10**, 477–91.
- Nadel J (2002). Imitation and imitation recognition: functional use in preverbal infants and nonverbal children with autism. In AN Meltzoff and W Prinz, eds. *The Imitative Mind: development, evolution, and brain bases*, pp. 42–62. Cambridge, Cambridge University Press.
- Parkinson B (2005). Do facial movements express emotions or communicate motives? *Personality and Social Psychology Review*, **9**, 278–311.
- Pollick AS and de Waal FBM (2007). Ape gestures and language evolution. *Proceedings of the National Academy of Science*, **104**, 8184–9.
- Rendell D, Seyfarth RM, Cheney D, and Owren MJ (1999). The meaning and function of grunt variants in baboons. *Animal Behaviour*, **57**, 583–92.
- Rendell LE and Whitehead H (2005). Spatial and temporal variation in sperm whale coda vocalisations: stable usage and local dialects. *Animal Behaviour*, **70**, 191–8.
- Ristau C (1991). Aspects of the cognitive ethology of an injury-feigning bird, the piping plover. In C Ristau, ed. *Cognitive Ethology*, pp. 91–126. Hillsdale, NJ, Erlbaum.
- Rizzolatti G (2005). The mirror neuron system and imitation. In S Hurley and N Chater, eds. *Perspectives on Imitation: from neuroscience to social science*, Vol. 1, pp. 55–76. Cambridge, MA, MIT Press.
- Schaller GB (1976). *The Serengeti Lion: a study of predator-prey relations*. Chicago, University of Chicago Press.
- Seyfarth RM and Cheney DL (2003). Signalers and receivers in animal communication. *Annual Review of Psychology*, **54**, 145–73.
- Striano T, Henning A, and Stahl D (2005). Sensitivity to social contingencies between 1 and 3 months of age. *Developmental Science*, **8**, 509–18.
- Tinbergen N (1964). The evolution of signaling devices. In W Etkin, ed. *Social Behavior and Organization Among Vertebrates*. Chicago, University of Chicago Press.
- Tomasello M (1998). Intending that others jointly attend. *Pragmatics and Cognition*, **6**, 229–44.
- Tomasello M (2003). *Constructing a Language: a usage-based theory of language acquisition*. Cambridge, MA, Harvard University Press.
- Tomasello M, Call J, Nagel K, Olguin R, and Carpenter M (1994). The learning and use of gestural signals by young chimpanzees: a trans-generational study. *Primates*, **35**, 137–54.
- Tomasello M, Carpenter M, Call J, Behne T, and Moll H (2005a). Understanding and sharing intentions: the origins of cultural cognition. *Behavioral and Brain Sciences*, **28**, 675–709.
- Tomasello M, Carpenter M, Call J, Behne T, and Moll H (2005b). In search of the uniquely human. *Behavioral and Brain Sciences*, **28**, 721–7.
- Trevarthen C and Hubley P (1978). Secondary intersubjectivity: confidence, confiding and acts of meaning in the first year. In A Lock, ed. *Action, Gesture and Symbol*, pp. 183–229. London, Academic Press.
- Trivers RL (1971). The evolution of reciprocal altruism. *Quarterly Review of Biology*, **46**, 35–57.
- Trivers RL and Hare H (1976). Haplodiploidy and the evolution of the social insects. *Science*, **191**, 249–63.
- Turing AM (1950). Computing machinery and intelligence. *Mind*, **59**, 433–560.
- von Frisch K (1953). *The Dancing Bees: an account of the life and senses of the honey bee*. New York, Harcourt Brace Jovanovich.
- Wiener N (1948). *Cybernetics: or control and communication in the animal and the machine*. New York, The Technology Press.

Wilson EO (1975). *Sociobiology: the new synthesis*. Cambridge, MA, Harvard University Press.

Wilson M and Knoblich G (2005). The case for motor involvement in perceiving conspecifics. *Psychological Bulletin*, **131**, 460–73.

Wolpert DM, Doya K, and Kawato M (2003). A unifying computational framework for motor control and social interaction. *Philosophical Transactions of the Royal Society of London: B Biological Science*, **358**, 593–602.

Zuberbühler K (2006). Language evolution: the origin of meaning in primates *Current Biology*, **16**, R123–5.