CHAPTER 4

The Other Otherwise

> In equivalence to Watzlawick's statement that "one cannot not communicate" it has been found that also in human-robot interactions one cannot be not emotional.
>
> FRANK HEGEL

Affective Loops and Human-Robot Coordination Mechanisms

The unstable, indeed untenable, character of the distinction between the so-called internal and external aspects of the emotions indicates that the problem of affective behavior must be analyzed in some other fashion. Together with the emergence of the interactive approach, and its paradoxical reversion to the classical view of emotions as private inner events whose expression is secondary and contingent, it suggests an alternative.

Although it lies outside mainstream thinking in philosophy and the sciences, this rival perspective nonetheless has a long history; indeed, one finds an early version of it in Hobbes's analysis of the passions.[1] In recent years, it has acquired important allies in cognitive science, especially within the embodied cognitive science, and particularly in the neuroscientific research on mirror neurons and related mechanisms.[2]

The alternative approach stresses the interactive character of emotions and affective processes and sees affective expression as the core of an intraspecific coordination mechanism in which emotions constitute salient moments. The central hypothesis is that affective

expression structures an intersubjective dynamic through which agents mutually determine their emotional states and coordinate their intentions to act. Every one of us participates from birth in this dynamic, which acts at all levels of personal cognitive organization and influences not only our interactions with one another, as individuals and as members of groups, but also our interactions with the environment. More particularly, this affective dynamic shapes our social environment. According to the affective coordination hypothesis, the affective domain in the broadest sense is the site of our primordial, purely biological sociability, which precedes and encompasses the social relations studied by sociology and anthropology.[3] This primordial sociability forms a part of the basis on which cultural, ritual, and social rules of behavior (including kinship rules) are erected and, at the same time, constantly furnishes us with opportunities for deviating from these same rules.

From the point of view of affective coordination, the question of how we acquire knowledge of others' emotions assumes an altogether different cast than the one it has in the classical conception adopted by the computational approach in cognitive science and in the various moderate versions of the embodied mind thesis that are more or less openly accepted in social robotics research today. In the classical view, knowledge of others' emotions, just like that of others' minds in the Cartesian understanding, proceeds indirectly, through inference or by analogy. As a consequence, being able to recognize emotions is crucially important. The affective coordination hypothesis, in contrast, rejects not only the idea that access to others' emotions is obtained by analysis or simulation of expressive behavior, but also that such access rests ultimately on recognition.

It is characteristic of studies of the ability to recognize others' emotions—in the work, for example, of Ekman and Friesen[4] and Izard[5]—that the question of whether one recognizes a particular emotion can be given only one right answer, because recognition is

considered to be equivalent to correctly identifying others' emotions. Recognizing affective expression in another person is a matter, then, of determining whether one is dealing with anger, fear, joy, or disgust, for example. To this question there is only one right answer: anger in the case of anger, fear in the case of fear, joy in the case of joy, and so on. Any other response is erroneous and false—evidence of a failure of recognition. When one looks at the matter in terms of affective coordination, on the other hand, it takes on a completely different aspect. I can respond to another's anger by fear, by shame, by anger, or even by laughter. None of these responses, or, if you prefer, none of these reactions, is a priori a false answer. Any of them can permit two agents to coordinate their behavior. Faced with your anger, my shame does not constitute an error, but a coordination "strategy" that may very well lead to an equilibrium.[6] Recognizing assumes a prior act of cognizing—that is, of knowing, in the sense of discovering a particular state of affairs. Reacting is an acting in return, a response, that transforms the meaning and the consequences of the original action. To be successful, a response does not require that a normative criterion concerning the proper interpretation of the original act be satisfied, least of all when a person's reaction acts upon and modifies the affective state of another.

Certain recent results, now securely established in neurophysiology with the discovery and subsequent study of mirror neurons and mirror systems, have given us a good idea of how reciprocal affective action works, even if they do not wholly explain it. Mirror mechanisms produce neuronal coactivation during the course of interaction between someone who acts and someone who observes him. The same neurons are aroused in the observer as the ones that in the observed agent are responsible for the performance of an action or for the display of an affective expression. Vittorio Gallese has proposed that, in the latter case, these mirror phenomena be regarded as embodied mechanisms providing access to the emotions

of another person. The access made possible by such mechanisms amounts to a kind of understanding, achieved through a process that Gallese calls "attunement,"[7] which allows the observer to participate in the emotional experience of the observed at an unconscious, subpersonal level by sharing its neurological underpinnings. Gallese considers this form of *embodied* emotional participation to be a basic manifestation of empathy, which he defines as a dynamic interindividual coupling between two or more agents in sensorimotor interaction that suspends the boundary between oneself and another by means of neuronal coactivation. He points to experimental evidence suggesting that the neuronal coactivation produced by mirror mechanisms temporarily prevents the nervous systems of people interacting with one another from being able to tell who is the agent and who is the observer of the affective expression, until sensory feedback takes place. Coactivation, in other words, causes intraindividual space and interindividual space momentarily to coincide.[8]

Our emotions and our empathic reactions are neither private productions nor solitary undertakings. They are joint enterprises in which two or more people take part. Emotions and affect do refer to the body, but this body is not limited to an individual organism; instead, emotions are embodied in what may be called a "social body." From this point of view, the organismic embodiment central to internal robotics, which seeks to unite cognitive processes and emotional bioregulation, meets up with the interindividual embodiment of expression that the external robotics of emotion exploits. In the affective coordination approach, the inner and outer aspects of emotion are not merely juxtaposed, as they are in the affective loop approach described in Chapter 3. They actually combine and merge with each other, for the outer affective expression of one agent is directly responsible for the inner reaction of the other.[9]

Affective expression is a means of direct and reciprocal influence among agents that, even if it is situated at the subpersonal level and

supposes a transient erasure of the boundary between self and other, does not cause the identity of the agents in communication with each other to be extinguished. To the contrary, it is owing to affective coordination that the agents' identities are constantly being established, strengthened, and redefined.[10] This influence is direct to the extent that it neither requires nor assumes any intermediary between the one's affective expression and the other's neuronal reaction. The reaction flows solely from the perception of affective expression. It requires no rational calculation or theorizing that is supposed to give one agent indirect access to the inner and private affective states of the other. The influence exerted by affective coordination is reciprocal as well, because the other reacts immediately and expressively to my affective expression, and his reaction brings into existence in my brain a new neural state. Affective coordination thus leads to affective codetermination—an affective loop in the strong sense—that shapes agents' affective expression and their intentions to act.

It is this direct affective influence that the artificial agents designed by external robotics seek to exploit and on which their success as social robots depends. Human beings are caught up from birth in an intersubjective dynamic by which affects are reciprocally determined. Social robots, though they have no mirror mechanisms capable of directly putting them in touch with our emotional experience, are now being introduced into this dynamic. They will need to find a place, a little as our pet animals do, in the affective dialogue that is constantly taking place among us. Unlike our pets, however, social robots have so far found it difficult to fit in. Their limitations arise in large part from the fact that they are not capable of reacting appropriately to the affective expression of human partners. The problem is not just the rudimentary nature of their affective response, but also, and still more significantly, an inability to coordinate their responses with those of their partners.

From this point of view, the ability to artificially produce emotions and empathy does not depend on having a "good" model or a

"deep" model of the physiology of natural affects, whether animal or human. It depends instead on the ability of robots to recreate, in the course of interaction with human beings, certain fundamental aspects of the phenomenology of interindividual affective coordination. It depends, in other words, on their ability to include both their human partners and themselves in a recursive interaction process influencing the emotions of human partners and aligning their disposition to act with a corresponding disposition in robots. If robotic agents capable of functioning smoothly as part of this dynamic can be created, it will bring about a social process of human-robot coevolution.

The moral (and political) issues raised by such a process are wholly separate from the question whether robots' emotions are true or authentic. Living with emotive and empathic robots will amount to sharing with them an affective experience that is more or less similar to the one we have in our relationships with pets or that a child has with a stuffed toy animal. These relationships are not artificial. Nor are they false, though they may very well be unbalanced, confused, or perverse. Nevertheless, whether they are altogether healthy, in neither case is it a question of being fooled by our dog or our teddy bear. Childless people who leave all their money to their cats are not really victims of feline cunning. Of course not, it will be conceded—but an intelligent artificial agent can fool us, even if a stuffed animal or a real animal cannot. It is far from clear, by the way, that real animals cannot fool us;[11] but as far as intelligent artificial agents are concerned, the ability to deceive has nothing to do with the presence or absence of an affective dimension. A robot can lead its interlocutors into error by giving them false information, something that it can very easily be programmed to do.

Classically, a concern with the truth or authenticity of an emotion locates the ethical dimension of our affective relationships in interiority and, more precisely, in intention. The sincerity of an intention

is considered to be the criterion of an emotion's truth or authenticity.[12] This assumption has no meaning in relation to social robotics, because robots feel nothing. The problem is not that their intentions are insincere, but that they do not have any. To dispose of this difficulty, internal robotics looks to create artificial agents whose behavior is guided by mechanisms that resemble those that are thought to cause, or else in one way or another to accompany, an "inner feeling." By resorting to models of human physiology, in other words, it hopes to be able to make up for the lack of emotion that condemns robots to a life of pretense—to having only feigned and false emotions. The affective coordination approach posits, to the contrary, that the emotions of an agent, whether natural or artificial, are interactive, distributed phenomena, in which two or more interacting agents participate. Accordingly, the ethical dimension is not located in either an internal or an external domain, but in a dynamic that operates on, and transforms, the very agents whose behavior constitutes and sustains it.

Radical Embodiment and the Future of the Social Robotics of Emotion

The idea that the nature of affective relations needs to be reconsidered, and the dynamic of emotions and empathy analyzed in a novel way, has found support in the so-called radical embodiment approach in philosophy of mind and cognitive science. Radical embodiment rejects the extended mind hypothesis, even where its "extension" includes social and intersubjective factors among the external resources on which an agent draws in order to carry out a cognitive task.[13] The radical approach to embodiment proposes a much more revolutionary redefinition of the borders of the mind, freeing it from the spatial dimension within which the debate over extended cognition has confined it up until now.[14]

Radical embodiment, particularly in the enactive version originally developed by Francisco Varela,[15] distinguishes itself not only from the classically dominant tendency in cognitive science, but also from the moderate versions of the embodied mind thesis that seat it in the brain and then extend it, in an ad hoc fashion, outside the intraindividual space. Enaction holds that the mind is situated in—or, rather, arises from—the complex regulative dynamic through which the agent's nervous system couples her body and her environment and thus makes cognition and knowledge of environmental context and of others possible. The radically embodied mind is not a spatial entity in the Cartesian sense of a *res extensa*. It is the result of a dynamic coupling that is irreducible to the classical alternative, inherited from Descartes, between an unextended immaterial substance and extended matter. Because it emerges from a process of reciprocal specification that connects the agent's nervous system with her body and her environment, the mind escapes the spatial framework that any disjunction between internal and external, or between organism and environment, cannot help but assume. The mind emerges, in other words, from a process of coevolution whose imbricated structure inevitably locates mind in the world, and vice versa.[16]

The radical embodiment approach involves more than merely adopting an abstract and speculative theoretical stance at odds with the classical computational conception of a "naked mind" that would be identically implementable in very different materials—as long as a certain "functional equivalence" is preserved between these various "realizations" of the mind. The sheer unreality of the computational conception will be apparent if one considers what a naked mind really implies: take away the body, the environment, and other agents, and all cognitive processes inevitably come screeching to a halt. Modeling and exploring cognitive phenomena become impossible. Contemporary synthetic models of artificial agents, based on the inseparability of

brain, body, environment, and other cognitive systems with which agents interact, illustrate the fruitfulness of radical embodiment as a methodological principle. In contrast with a purely hypothetical naked mind, these agents can be designed and actually constructed as recursive systems that perpetually determine and modify one another's state.[17]

Radical embodiment does a better job than the extended mind hypothesis in explaining why Otto would have chosen, as surely he must have done, to get to the Museum of Modern Art by hailing a taxi and asking the driver to take him there. The mind is neither an item of personal property nor something that belongs to isolated individual agents. It is a process in which all agents jointly participate. We believe that extending this principle to the robotic modeling of emotions and empathy will lead to dramatic advances. A relational approach to affective processes, taken together with recent results in the study of mirror mechanisms as well as enactive modeling, will do more than associate the production of emotions with their expression. Not only will a relational approach annul the longstanding divorce of production from expression, it will restore their interdependence by bringing out the complex network of connections that bind together agents through affective processes. In a single stroke, rejoining them does away with all the dichotomies that have been relied on for so long to describe and analyze empathy and emotion: between inner and outer, between private and public, between personal and social, and, not least of all, between true and false.

According to the classical approach, an affective dynamic results from private, internal generative processes that sometimes give rise to an external expression that is both public and social. This expression is supposed to be analyzed subsequently by other agents who are capable of activating in their turn generative processes of the same type, once they have recognized the emotion that has been expressed. According to the affective coordination approach, by contrast, an

affective dynamic is a process through which agents mutually transform themselves. They act upon one another on various levels, not only affective and cognitive, but also physiological ("He made me so angry, my stomach hurt").[18] The generation and expression of emotions are complementary and interchangeable moments of the dynamic of affective coordination. They cannot be entirely separated from each other, for affective expression by one person is responsible in part for generating emotion in the other.

Such a perspective deconstructs the true / false opposition by showing the emptiness of the idea that robots will manifest "authentic" emotions—and will not deceive human beings—only when their affective expression proceeds from an internal architecture that reflects, at some suitable level of abstraction, individual animal and / or human affective processes.[19] In effect, then, the relational perspective urges social robotics to explicitly formulate and complete the paradigm shift it is presently undergoing—a shift whose signs we detected earlier, having noticed that research in social robotics constantly and inevitably violates the theoretical distinction between the internal and external aspects of emotions.

The relational perspective proposes a synthetic approach in which creating "robotic emotions" and "robotic empathy" means equipping robots with *human-robot affective coordination mechanisms.* This requires, to begin with, that we no longer look to build robots that artificially reproduce human or animal emotions, conceived as properties of individual agents. We need instead to place the intersubjective dynamic of affective coordination at the center of research on emotion and empathy and develop mechanisms that will make it possible to construct robots that are *essentially interactive affective agents.* These mechanisms, and the emotive and empathic processes they generate, are closely linked with the specific characteristics of the interaction dynamic in which agents are engaged. It is, we believe, on this point—the creation of artificial emotions and empathy as moments

of a dynamic of human-robot transformation that coordinates the behavioral dispositions of interacting human and artificial agents, rather than as inner computational or physiological processes that are capable of being given outward expression—that current research has already begun to converge.[20] Social robotics ought therefore to openly declare this to be its aim.

What are the most promising robotic platforms, the architectures best suited to creating artificial agents that will be truly interactive affective agents? Only the coevolution of humans and robots will eventually be able to tell us. Robotic agents of the present generation nonetheless give us a glimpse of what interaction with human beings in one way or another may one day look like. Until now, however, robots have engaged with people on an emotional level only to a very modest extent. Here are three rather well-known robots—Geminoid, Paro, and KASPAR—that illustrate some of the difficulties that attempts to establish a robust human-robot affective dynamic encounter, as well as the limitations of what has been achieved so far.

Geminoid: Social Presence, or Acting at a Distance

Geminoid, the creation of Hiroshi Ishiguro at Osaka University, is an android robot whose appearance almost perfectly reproduces that of its designer. Outwardly, then, Geminoid is Ishiguro's double. Nevertheless, it is not an autonomous robot, capable of acting or moving around by itself. It is a twin, as its name indicates ("Geminoid" is derived from the Latin *geminus*), but only in a very attenuated sense, for its abilities fall far short of those of ordinary mortals, to say nothing of a gifted scientist. Geminoid is essentially a doll, an extraordinarily sophisticated marionette—but a machine all the same, bolted to its chair, attached by a series of cables to a control room, and pumped up by a pneumatic system to look healthy and fit. Geminoid is capable only of moving its head, eyes, mouth, and facial muscles. Nevertheless, it can see and speak. It can also hear what is said to it and carry on a normal conversation. Yet it cannot manage these feats by itself or, more precisely, by itself alone.

Geminoid is remotely controlled, with the help of a computer, by an operator who sees and hears what the robot "sees" and "hears." It is the operator who responds as well. The robot should, in principle, be able not only to transmit the operator's words, but also to reproduce his facial expressions and mouth movements. The operator constitutes the robot's soul, in the altogether classical sense of something that is introduced from outside into an agent's body and that animates it. This arrangement makes it possible for Ishiguro to be traveling, say, in Moscow, and at the same time be present at a meeting of his laboratory staff at ATR on the outskirts of Kyoto.[21] Thanks to Geminoid, Ishiguro can act and react somewhere he is not. In a certain sense, he can be physically present in two places at the same time. As a practical matter, however, Ishiguro's mechanical body is inhabited by whoever—teacher, student, researcher—is sitting at the console in the control room. In addition to making possible what might be called three-dimensional teleconferencing, Geminoid is a tool for exploring the uncanny valley. Endowed with a physical appearance as similar as possible to that of its creator, and remotely controlled by a human operator whose intellectual abilities and capacity for social communication equal or exceed those of an average person, Geminoid ought to allow us to determine more precisely what causes the mysterious uneasiness we feel in the uncanny valley and to have a better understanding of what exactly relations between robots and human beings involve.

Zaven Paré and Ilona Straub conducted a detailed series of communication experiments with Geminoid over the course of three weeks, with the two of them taking turns questioning the robot and operating its controls, Paré sitting in front of the robot while Straub was stationed in the control room, and vice versa.[22] What these

experiments demonstrate is the reality of "action at a distance." This kind of action has to do not merely with the fact that Geminoid permits the person commanding it to act in a place where he or she is not physically present. Geminoid makes action at a distance possible in the sense we considered earlier in this chapter with regard to affective coordination, namely, action that takes place at the subpersonal level and that two or more agents take part in.

The experiments are notable chiefly for the robot's stubborn insistence on doing just as it pleases. In reality, Geminoid cannot do much. And owing to bugs, interference, technical difficulties, control problems, and gaps in speech reminiscent of a poorly dubbed film, it has a very hard time doing the few things it is able to do. As a result of all these shortcomings, Geminoid constantly interposes itself between its operator and its interlocutor. Nevertheless, the effect of being there—the robot's social presence—comes through. This effect may be defined as an "action" that the robot exerts on its interlocutor by its presence alone. Anyone who converses with Geminoid has the impression of being in the presence of another person. This ability that it has of acting on us, its human interlocutors, amounts to action at a distance to the extent that it is something that it does to us and that, paradoxically, it does to us by doing nothing, by simply being there.

Knowing that a robot such as Geminoid sees, noticing that it looks, and seeing that *it's looking at me* are not the same things. In the last case, looking at me is something that the robot does to me. "I," the object of its gaze, am sensitive to this gaze. It does something to me. In doing this something, the robot, without moving from where it is, acts on me. It disturbs me, reassures me, worries me.[23] It will be objected, of course, that the robot does not really act on me. On the contrary, it is because I see and understand that the robot is looking at me that I am disturbed. There is therefore no action at a distance here, only an awareness of being the object of another's gaze, of being

the target of it, and a knowledge, perhaps innate, of the possible consequences of this state of affairs.

Yet anyone who has even the slightest experience interacting with social robots knows that the robot does not "look" at us in the relevant sense of the term. It shows no interest in us; often, in fact, it sees nothing at all. This in no way changes, or lessens, the impression we have of being the object of another's action. It is true that in this case the interlocutor knew that it was really another person, in the control room, who saw him (or her), but a crucial aspect of these experiments is Geminoid's obstinacy in interposing itself between the two experimenters. Even when the robot's behavior eluded the operator's control, the robot continued to assert its presence, and indeed, in a certain sense, it asserted it still more forcefully.

To be sure, it is possible to say that here the feeling of being disturbed, for example, was unjustified, that it was an error—possibly an inevitable one, Mother Evolution having made us in such a way that we know how to recognize, by certain signs that roboticists exploit, when we are the objects of another's gaze. This ability evolved because being observed by another organism is often a biologically important situation (for example, to be looked upon as a possible sexual partner by another member of one's species)—and sometimes a dangerous one (to be looked upon as prey by a predator).

Yet while knowing that an impression is false changes nothing with regard to the impression itself, as in the case of some optical illusions,[24] we seem no longer to be dealing with knowledge in this case—I do not need to *know*, strictly speaking, that another agent is looking at me—but with something that is rather nearer to a reflex. If this is so, then it is no longer necessary to imagine some intermediary, a mental representation, for example, between another's gaze and my reaction. Even if one assumes the existence of a module, itself representational, that is responsible for detecting the other's gaze, the functioning of this module (by hypothesis, a module in the sense

this term has in philosophy of mind) is entirely impenetrable to me. I have access only to its results, which depend on the other's action and which, in this sense, are under its control. Therefore, it is indeed the robot's gaze that acts upon me. In commanding the module's result, it leads to my reaction, typically described as spontaneous, and produces in me a feeling of social presence. The robot acts directly on me from where it is seated no less surely than the physician's reflex hammer does when he taps my knee with it.

The idea of action at a distance seems very strange to us today. Traditionally, it is associated with magic, and since Descartes, the modern scientific view of the world has rejected it. Yet the discovery of mirror neurons explains how such an action at a distance between two agents may be possible. Robotics demystifies it, tames it, makes it intelligible. Geminoid shows that there is no mysticism at work here; acting at a distance corresponds neither to an ineffable sense of there being something "between us," nor to a mere anthropomorphic projection. The presence of another is a phenomenon that can be contrived and implemented with the aid of a machine and that therefore can be reproduced and analyzed. The experiments performed by Paré and Straub cast light on what we find unsettling, disconcerting, and on what makes it difficult for us to recognize this way of acting for what it is. The remote effect of a robot's mere presence, without it having to do anything at all, is yet, paradoxically, an action—an action without either act or actor, as it were! Ordinary language may be confusing here, but we have at our disposal an adequate scientific vocabulary. If this action may reasonably be said to be without either act or action, it is because it takes place and unfolds at a subpersonal level where the difference between self and other is not clearly established. It is nevertheless something that the robot does to *me,* to the extent that *I* passively experience a presence of which it is the cause.

Action at a distance by a robot helps us see that the effect we have upon one another, by the simple fact of our presence, takes place at a

subpersonal level. It takes place without our being able to attribute an action to *someone* or to recognize as the author of the action what we discover to be the cause of the experienced effect. The robot, a machine that we ourselves have constructed, does not act by magic. But we are far from knowing exactly how it does what it does, and we are still farther from being able to determine exactly what effect the robot has on us. Unease, discomfort, familiarity, the sense of being kept at a respectful distance—all these feelings are somehow in play, without our being able to locate the effect definitively in one or another register. The exploration of the uncanny valley made possible by Ishiguro's robot is aimed at answering these questions.

Now the essential thing that Geminoid lacks, considered both as a robot, when it escapes the operator's control, and as part of the team that robot and operator normally form, is the ability to react to the presence of another on a level where it is capable of making its presence felt. There are at least two reasons for this. First, Geminoid is a communication interface that, in a sense, is supposed to stay in the background, behind the operator, whom it allows to be present somewhere he or she is not. Second, the operator experiences the presence of his or her interlocutor only as an image on a screen. Unlike the robot's human partner, the operator does not physically experience the social presence of the other person with whom the robot interacts. No affective loop can be established.

Paro, or Proximity: A Return to Animal-Machines

Paro is described by its creator, Takanori Shibata, as a "mental assist robot" designed to interact physically with human beings.[25] It is utilized chiefly as an animal companion for therapeutic purposes in hospitals and homes for the elderly. Paro has the appearance of a baby harp seal (*Pagophilus groenlandicus*) and weighs 2.8 kilos (a little more than six pounds). Like Geminoid, it has no mobility: it is incapable