VI. FIRST-PERSON AND THIRD-PERSON MINDREADING

PIERRE JACOB

1. INTRODUCTION

My topic is what the developmental cognitive psychologist Simon Baron-Cohen (1995) aptly calls "mindreading". As the label suggests, mindreading is the human ability to recognize the presence of minds. Full-blown human mindreading culminates in the practice of representing and ascribing states of mind – states of minds of others and states of one's own mind. Typically, human beings ascribe emotions and socalled propositional attitudes (such as intentions, beliefs and desires) to other agents for the purpose of explaining and predicting their intentional actions. A person represents her own beliefs, intentions and desires when e.g., she is in the process of revising her goals and/or her beliefs and is checking the consistency of her own mind. Alteration of mindreading can result in "mindblindness". As Baron-Cohen (1995) has put it, "imagine what your world would be like if you were aware of physical things but were (mind-) blind to the existence of mental things [...] blind to things like thoughts, beliefs, knowledge, desires, and intentions, which for most of us self-evidently underlie behavior."

If and when it is reliable, first-person mindreading – traditionally called "introspection" – gives rise to self-knowledge or true beliefs about the contents of one's own mind. Third-person mindreading gives rise to knowledge of, or true beliefs about, the contents of other minds. From an evolutionary standpoint, the adaptive benefits of third-person mindreading among humans are quite obvious. If one individual is contemplating the choice between cooperation and competition with another agent, then it is useful for the former to be able to detect accurately the latter's goals and intentions and vice-versa. Arguably, if one did not know the content of one's *own* mind, then one could not even contemplate the choice between cooperation and competition. If so, then the adaptive advantages of both third-person and first-person mindreading stand or fall together. Plausibly then, they must have co-evolved in tandem.

The goal of my paper is to assess a recent proposal in the cognitive sciences and in the philosophy of mind according to which all aspects of mindreading can be accounted for by a simple cognitive heuristic called

"mental simulation". I will argue that some fundamental aspects of fullblown third-person mindreading lie outside the scope of mental simulation and that mental simulation fails to account for first-person mindreading. Before I proceed, however, I will first make a couple of assumptions explicit and then I will make three disclaimers.

2. A PAIR OF ASSUMPTIONS

Human beings are self-conscious and they are conscious of others. I assume that both one's awareness of others and one's self-awareness can be divided into awareness of two kinds of properties: awareness of non-psychological or bodily properties and awareness of psychological properties. There are two broad ways in which one can become aware of both the bodily properties of others and one's own bodily properties.

On the one hand, one can use one's sense organs to perceive the non-psychological properties of others. One can also see, hear and smell some – not all – of one's bodily parts. One can hear one's own voice. One can have proprioceptive information about the relative positions of one's limbs and about one's posture. Indeed, as Gibson (1979) insisted and as contemporary cognitive neuroscience confirms, the visual system could not represent some of the properties of distal stimuli unless it took into account information about some of the observer's movements. For example, the visual computation of the position of an object relative to an agent is based on information about the agent's eye- and head-movements. Thus, visual percepts of non-mental objects may depend upon the processing of non-perceptual information about the bodily self or what Bermudez (1998) calls "the embodied self".

On the other hand, one can also come to learn facts about human anatomy and human physiology, and thus about one's bodily properties, not by perceiving bodily properties — neither the bodily properties of others, nor one's own bodily properties —, but by reading books about them. By reading books, one can come to know anatomical and physiological facts about e.g., the human skeleton, the human digestive system, the human cardiovascular system, the human immune system or the human nervous system. In so doing, one becomes *conceptually* aware of the bodily (or non-psychological) properties of both oneself and others.

Psychological self-knowledge (or first-person mindreading) raises two questions. First, I can be aware both of my non-psychological properties and of my psychological properties. Given that I can be aware of my non-psychological properties in two ways, the first question is whether my awareness of my own psychological properties comes in two varieties too. Can one be both perceptually and conceptually aware of one's own psychological properties? Secondly, I can be aware of both your psychological properties and my psychological properties. I can do third-person mindreading and first-person mindreading. It is obvious that I must rely on the observation of your behavior in order to know your mind. What about first-person mindreading? Do I also need to observe my own behavior to do first-person mindreading?

The word "introspection" refers to the process whereby one person is aware of her own psychological properties. Many philosophers – especially in the classical empiricist tradition – have assumed that one can be perceptually aware of one's own psychological properties. I think that this is a deep mistake. Observation of the behavior of others is required to know their psychological properties. One can be perceptually aware of some of one's non-psychological properties. But awareness of one's own psychological properties can only be conceptual. Introspection is a conceptual, not a perceptual, mechanism (as I will argue). On my view, Shoemaker's (1968) well-known claim that one's knowledge of oneself is immune to errors of misidentification is restricted to introspective knowledge of one's own psychological properties.

3. THREE DISCLAIMERS

My first disclaimer has to do with the scope of the concept of knowledge as I understand it here for the purpose of analyzing how humans come to know the contents of either their own minds or the minds of others. Philosophers – and especially contemporary analytic epistemologists – are concerned with a restricted sense of the English verb 'to know'. In this restricted sense, a person can be said to know proposition p only if she believes that proposition p is true, proposition p is indeed true and some further condition is satisfied so that the person's belief has not been formed out of shere luck or mere fantasy. According to externalist epistemologists, the belief must have arisen by virtue of some reliable mechanism. According to internalist epistemologists, a person will fail to know that p unless she has a true justified higher-order belief

¹ A claim anticipated by Wittgenstein and explored in his own way by Evans (1982).

that her true first-order belief is itself justified. In this paper, I do not use the verb "to know" in this restricted epistemological sense. Like psychologists, if and when I say that person X knows either that she believes so and so or that someone else desires such and such, I do not mean that X's knowledge qualifies as "knowledge" in the epistemologist's favored sense.

Secondly, one should not expect from the cognitive scientific study of third-person mindreading a response to the challenge of philosophical *skepticism* about knowledge of other minds. Philosophical skeptics worry about whether knowledge e.g., of the external world is possible *at all*.² Suppose I claim that I know that my neighbor Ann is at home because I have seen her car parked in front of her house. As a matter of fact, Ann's car would not be parked in front of her house unless she was at home and I know it. Now, seeing Ann's car parked in front of her house may well be a reliable method for knowing whether she is at home. But the philosophical skeptic will not be satisfied, because the explanation of how I know something about Ann (namely that she is at home) takes for granted that I know something else about her car (namely that it is parked in front of her house). And the skeptic refuses to take any knowledge for granted.

Similarly, in the case of knowledge of other minds. Suppose I claim to know what Ann thinks because either I asked her or I asked her husband. Asking her or asking her husband may be a reliable method for knowing what Ann thinks on a particular topic. However, if this is the way I come to know what Ann thinks, then I take for granted either that she knows what she thinks or that her husband does. Therefore, I fail to satisfy the skeptic's requirement for a perfectly general account of knowledge of what someone thinks – i.e., an account that does not presuppose any other knowledge. So my second disclaimer is that we should not expect the cognitive scientific study of third-person mindreading to meet the challenge of skepticism about knowledge of other minds. Nor should we expect it to offer a perfectly general account of knowledge of other minds. Rather, we should expect it to tell us something about the actual mechanisms underlying our knowledge of other minds

Thirdly, the idea that *introspective* beliefs about facts involving one's own psychological properties (such as believing so-and-so or experiencing such-and-such) enjoy a unique *epistemic authority* or privi-

² See Stroud (1989).

lege has played a fundamental though different role in traditional rationalist epistemology and in traditional empiricist epistemology. In rationalist epistemology, the primary target of introspective self-knowledge are one's own thoughts. In empiricist epistemology, the primary target of introspective self-knowledge are one's own sense-data or perceptual experiences. I do not think that the traditional epistemological privileges ascribed to introspective self-knowledge are justified. I do not take for granted either that self-knowledge is a priori knowledge or that it is infallible.

On the one hand, in rationalist epistemology, psychological selfknowledge was taken to be one of two paradigms of a priori human knowledge, the other paradigm being knowledge of mathematical truths. Rationalist epistemology has three ingredients. First, it is of the essence of the Cartesian mind that it entertains or forms thoughts (as opposed to having e.g., perceptual experiences). Secondly, thoughts have concepts or ideas as constituents. Thirdly, the mind is transparent to itself: one cannot have a thought of which one is not aware. In rationalist epistemology, what secures both the priority and infallibility of one's introspective awareness of one's own psychological properties is that all the psychological properties a mind can exemplify are properties of thoughts or judgments, not experiences. Given that the mind is transparent to itself or that thoughts are reflexive in the sense (of the transparency assumption) that one cannot entertain a thought (or make a judgment) unless one is aware that one is, it follows that introspective knowledge of one's own mind is a priori and, according to Descartes, infallible.

In traditional empiricist epistemology, on the other hand, the most primitive and elementary constituents of minds are perceptual experiences or sense-data, not concepts. Concepts (of either psychological or non-psychological properties) are logical constructions out of sense-data. According to much traditional empiricist epistemology from Locke to Russell, knowledge of the external world – knowledge of mindindependent facts – is twice dependent on psychological self-knowledge. First of all, knowledge of mind-independent facts depends on the epistemologically antecedent knowledge of mental or psychological facts about oneself (such as that one is having a particular perceptual experience or sense-datum). Secondly, one's knowledge of mental or psychological facts about oneself derives in turn from one's direct quasi-perceptual acquaintance with some mental entity present to or in one's mind, i.e., the sense-datum or perceptual experience. If one's

awareness of one's sense-data consists in being acquainted with them, then one is made aware of one's sense-data by some kind of quasi-perceptual process or peering inside at one's own perceptual experiences.

Both the rationalist and the empiricist pictures of introspective self-knowledge have come under serious criticism in contemporary philosophy and psychology. On the one hand, the empiricist model of a quasi-perceptual process whereby one becomes self-aware of one's own perceptual experiences has been systematically dismantled. For detailed criticism of the quasi-perceptual model of psychological self-knowledge, see the last section of this paper.

On the other hand, the Cartesian picture of introspection raises at least two issues. First, the asymmetry between first-person and thirdperson mindreading that results from a Cartesian picture of introspective self-knowledge raises a genuine puzzle. As Davidson (1984, 1987) recognizes, the asymmetry between first-person and third-person mindreading takes it for granted both that claims to know one's own mind are made independently of any empirical evidence and that they enjoy an epistemic authority of which third-person claims to know the minds of others are deprived. The puzzle is: why should claims without evidential support have more epistemic authority than claims based on evidence? The second question is: given that the Cartesian assumption that the mind is transparent to itself has come under heavy attack, what is left of the Cartesian picture of the special epistemic authority of introspective beliefs? Since Freud, it is commonly accepted that one may be blind to some of one's own beliefs and desires. You may be better positioned than I am to get insight into some of my beliefs and desires. Furthermore, a human mind does not merely entertain thoughts and propositional attitudes; it also has perceptual experiences. On the Cartesian picture, the immunity to error of the mind's introspective beliefs about itself was secured by the joint assumptions that it is of the essence of the mind to entertain thoughts and that one cannot entertain a thought unless one knows that one is doing so. But how could such assumptions entail that one's introspective beliefs are exhaustive and exhaustively true a priori? How could they secure a priori true introspective beliefs about the content and attitude of each of one's own propositional attitudes – both one's occurrent propositional attitudes and one's dispositional propositional attitudes? Furthermore, how could such assumptions secure introspective a priori knowledge about one's perceptual experiences at all? At the minimum, the Cartesian picture of self-knowledge would

need to assume that perceptual experiences are judgments (or that the phenomenal character of a perceptual experience is derivative from a judgment)³.

4. WHAT IS MENTAL SIMULATION?

The idea of mental simulation involves three complementary ingredients. First, the idea of mental simulation as a basis for mindreading derives from the rejection of the picture of mindreading as a capacity for detached theorizing. Secondly, much of the appeal of mental simulation derives from the fact that it is at once a simple and a general-purpose, not a domain-specific, mental heuristic. As such, it holds the promise of unifying various aspects of human cognition. The idea is that any human cognitive mechanism can be used "off-line" in a "pretend" mode. Mental simulation is this pretend mode. Finally, proponents of the application of the mental simulation strategy to mindreading assume that the fundamental goal of mindreading is mental mimicry or the duplication of someone's mental life. The idea that mental simulation has the potential to account for the human mindreading capacity has arisen in two scientific contexts: in the context of the investigation by philosophers and developmental cognitive psychologists of the human capacity to do mindreading and in the context of the cognitive neuroscientific study of motor cognition.

The simulation approach to mindreading derives some of its appeal from some of the shortcomings of some of the rival approaches to mindreading which have emerged in both philosophy and cognitive science in the last twenty years or so. One such influential rival approach is the so-called "theory-theory" (or TT) approach to mindreading (associated with the work of Gopnik, Meltzoff and Wellman).⁴ This approach has arisen in response to two fundamental questions: 1. What is the fundamental structure of the human adult ability to mindread? 2. How is it acquired both from an ontogenetic and a phylogenetic point of view? According to TT approaches, the response to the first question is that human adults tacitly know a naive psychological theory comprised of naive psychological laws ranging over abstract unobservable mental states such as intentions, beliefs and desires. The response to the ontogenetic part of the second question is that human infants are "little sci-

³ A view that is highly questionable.

⁴ See Gopnik & Wellman (1992, 1994) and Gopnik & Meltzoff (1997).

entists": they use their science-forming capacity to discover the laws of naive psychology. In this respect, normal ontogenetic development undergoes a series of theory-changes pretty much like those exhibited in the history of scientific theorizing.

Some of the implausible features of TT views have prompted the emergence of an alternative nativist modular domain-specific approach. On this alternative modular view (associated with the work of Baron-Cohen and Leslie).⁵ mindreading results from a set of distinct special purpose modules.6 Each such module has a dedicated task: ID (intentionality detector), for example, takes patterns of motion of an object and ascribes goal or agency to the moving object. EDD (eve-direction detector) takes movements of the eyes in their orbits as input and finds some object as the target of the agent's gaze-direction. SAM, which is a shared or joint attention mechanism, takes as input the output of EDD and represents some object as the target of the joint attention of at least two agents (one of which can be the self). ToMM is the theory of mind mechanism proper, whose job is to deliver ascriptions of beliefs and desires to an agent.7 With an eye on the phylogenetic part of the second question, advocates of the nativist modular view entertain the hypothesis that several such modules may be present in the brain of non-human primates. In the brain of human infants, the outputs of ID and EDD serve as inputs to SAM whose output serves as input to ToMM. In the course of human development, this chain of modules unfolds according to a strict ontogenetic schedule.

Both theory-theorists and nativist modular theorists derive inspiration from the Chomskyan study of the language faculty where the ability of a speaker of a natural language to produce and understand an infinite set of grammatical strings of words of her language is explained in terms of her detached tacit knowledge of the grammar of the language. This is what simulation theorists reject and this is the first component of the mental simulation approach to mindreading. According to simulation theorists, mindreading reflects no so much a capacity for detached theorizing about mental states as a practical ability to *do* something, i.e., to

⁵ See Baron-Cohen (1995), Leslie (1994) and Leslie and Thaiss (1992).

⁶I take Carey & Spelke's (1994) and Spelke's (2000) favored notion of "core knowledge" to be more in line with what I call the "nativist modular domain-specific" approach to human cognitive development than with the "theory-theory" approach.

⁷ See in particular Baron-Cohen (1995).

engage into a certain kind of action, namely the impersonation or the replication of the mental lives of one's conspecifics.⁸ This is, I take it, what Gordon (1996, p. 11), a simulation theorist, has in mind, when he contrasts what he calls a "hot methodology" (one "which exploits one's own motivational and emotional resources") and a "cold methodology" (i.e., a "methodology that chiefly engages our intellectual processes").

The second and third features of the mental simulation approach to mindreading derive from the neuroscientific investigation of the motor cortex involved in the planning, the execution and the perception of human actions. The neuroscientific investigation of motor cognition in human and non-human primates has revealed that many of the brain mechanisms involved in performing particular actions are also involved in the mental preparation of (or in imagining) such actions and in the perception of such actions performed by others. On the one hand, most of the neural events whose unfolding gives rise to an *overt* (executed) action are involved either in its *covert* stages or in tasks of *motor imagery* in which the subject is required to imagine an action without executing it. On the other hand, most of the neural events involved in the covert stage of an action or in tasks of motor imagery are involved in the perception of a different token of the same type of action performed by another.⁹

One of the most startling findings of the cognitive neuroscience of motor cognition has been the discovery by Di Pellegrino et al. (1992), Rizzolatti et al. (1995), in area F5 of the pre-motor cortex of macaque monkeys, of a population of neurons – so-called 'mirror neurons' – that discharge selectively both when the animal performs some very specific object-directed hand movements and when the animal visually perceives the same kinds of object-directed movements performed by a conspecific. For example, mirror neurons have been found to discharge selectively for the execution and the perception of hand movements involved in holding, grasping, manipulating or tearing raisins or peanuts.

On the basis of such remarkable findings, simulation theorists have offered the suggestion that something like the mirror system in humans might be what Gallese (2003) calls a "shared manifold of interpersonal relations" (or social cognition). The idea is that mirror neurons offer a

⁸ I borrow the contrast between engaged (or immersed) cognitive activity and detached theorizing to Campbell (1994), who applied it to naive thinking about space and time.

⁹ Cf. Jeannerod (1997, 2001, 2002).

common neural basis not only for the execution and the preparation (or the imagining) of one's own actions, but also for the understanding of the actions of others. The idea that mental simulation is a cognitive heuristic underlying all forms of mindreading can be seen as a generalization of the discovery that in motor cognition, one and the same neural structure underlies the control and the imagination of one's own actions as well as the perception of the actions of others.

Any cognitive mechanism that takes standard inputs and produces standard outputs in response can be used off-line, i.e., it can be given "pretend" inputs and produce "pretend" output in response. For example, Currie (1995) argues that visual mental imagery is the off-line use of the visual system whereby it processes inputs from memory rather than retinal images. In imagining an action or in perceiving an action performed by another, one's motor motor system is used off-line. In the same vein, Stich & Nichols (1992, 1995) have hypothesized that one's practical reasoning (or decision) system can be used "off-line" in order to predict someone else's decisions. When used on-line, one's decision system takes one's beliefs and desires as inputs and produces decisions on the basis of which one acts. When used off-line, the same mechanism takes "pretend" beliefs and desires as input and produces a "pretend" decision, which is then used as a prediction of the other's decision. The hypothesis that any cognitive mechanism can be used off-line is the second ingredient of mental simulation.

The third ingredient of the idea of mental simulation is mental mimicry. As Gallese & Goldman (1998, p. 495) have emphasized, the motor system seems to "form a cortical system that matches observation and execution of motor actions." It is tempting – and indeed it has tempted many cognitive scientists and some philosophers as well – to suppose that the main function of the mirror system in both humans and nonhuman primates is to enable imitation. If, as most simulation theorists assume, the major function of the human mindreading system in turn consists in mental imitation (or in mimicking the mental life of a conspecific), then it will seem natural to assume that the mirror system underlies the human mindreading system. Both the view that the mirror system is a sufficient condition of human imitation and the view that the main goal of the human mindreading system is to reproduce the mental life of one's conspecific seem to me deeply mistaken. I shall start with the first mistake. ¹⁰

¹⁰ Note that the idea that the function of the mindreading system is to imitate

The view that the mirror system is a sufficient basis for human imitation embodies, I think, two mistakes. The first mistake is the conceptual confusion between imitation and mimicry. Imitation is very widespread in humans, but as Tomasello (2000) and Hauser et al. (2002) argue, monkeys and apes do not imitate very much, if at all. Spontaneous cross-modal imitation seems sharply limited to humans. There are mirror neurons in the premotor cortex of macaque monkeys, but macaque monkeys do not engage in imitation. By contrast, as observed by Meltzoff and Moore (1977, 1997), human infants only a few hours old are able to reproduce facial gestures that they have seen. Human infants are not bound to the mere reproduction (or copy) of bodily movements. Very early on, they can retrieve the intention underlying an action and perform a successful imitation of a failed exemplar of an action (Meltzoff, 1995). Human adults can imitate pretty much anything: human imitation ranges from non-intentional noises or movements to the intentional behavior of conspecifics. When imitating the intentional behavior of a conspecific, humans do not merely copy bodily gestures. They retrieve the intention underlying the imitated behavior.

Among members of non-human species including non-human primates, one finds ubiquitous instances of mimicry, not imitation. Whereas imitation is, mimicry is not, intentional behavior. Unlike human imitation, which may require the retrieveal of the intention underlying the observed piece of behavior, mimicry is a process whereby A's perception of some property exemplified in the environment automatically causes A to exemplify the property in question. For example, the monarchs are butterflies that are preys of birds (the bluejays). The former store and secrete a substance toxic to their predators. Other butterflies – the vice-roys – do not secrete the toxic substance, but instead they mimick the colors of the monarchs, thereby misleading the bluejays into 'believing' that they too are toxic. Mimicry is ubiquitous in nonhuman animals and in human behavior as well. Thus, emotional contagion, laughter and yawning are cases of mimicry whereby A's perception of B's behavior automatically (or non-intentionally) prompts A to reproduce B's behavior. But they are not instances of imitation.11

The second mistake is the assumption that the mirror system fires in response to the perception of social actions. Human and non-human

makes sense when applied to third-person mindreading, but not to first-person mindreading.

¹¹ Cf. Jacob & Jeannerod (2003, ch. 7).

primates can perceive the actions of their conspecifics. Some of the perceived actions are themselves directed towards conspecifics. Others are directed towards other things. Only actions directed towards conspecifics (e.g., communication) are distinctly social actions and only social actions are guided by social intentions, i.e., intentions to modify a conspecific's behavior.

When mirror neurons fire in the brain of monkey A observing the performance of an action executed by monkey B, B presumably intends to act on an object, but B does not intend ipso facto to modify A's behavior. B's intention is a motor, not a social, intention. The role of the mirror system in primates seems interestingly limited to matching execution and observation of *object*-oriented actions. Thus, mirror neurons are involved in the process of retrieving from perceptual cues motor intentions involved in an agent's object-directed actions. By contrast, neurons have been registered by Perrett et al. (1982, 1990) in a brain area called the superior temporal sulcus (STS), which is near but distinct from area F5 where the activity of mirror neurons has been recorded. Neurons in STS fire in response to the perception of movements of bodily parts that are not restricted to object-oriented actions and that are likely to convey emotional and social information about the agent. For example, they discharge upon seeing eye-gaze direction, movements of the lips, upwards and downwards head-movements, movements of the hand. But, as Rizzolatti et al. (2000, p. 548) emphasize, unlike mirror neurons, neurons in STS are purely perceptual, not sensori-motor, neurons. Thus, the most plausible interpretation of the mirror system in monkeys is that it constitutes an engaged proto-conceptual system designed for the recognition of object-oriented actions. It is engaged in the sense that it will not recognize an action unless the neural resources required for an animal to prepare the execution of that very action are active. Its recognitional powers are limited to object-oriented actions and motor intentions.

5. MENTAL SIMULATION AND THIRD-PERSON MINDREADING

There are two versions of the theory of mental simulation as it applies to third-person mindreading: Goldman's (1989, 1993, 2000) introspectionist version and Gordon's (1986, 1995, 1996) anti-introspectionist version. According to the former, knowledge of other minds derives from knowledge of one's own mind. In other words, I use

knowledge of my own mind as a model from which I inferentially construct knowledge of other minds. The introspectionist model of third-person mindreading is thus confronted to the preliminary question: how do I derive knowledge of my own mind? I will examine this question in the next and final section. I presently turn to the latter anti-introspectionist version of the simulation theory, according to which third-person mindreading consists in *pretence*. Both versions of the simulation theory, however, agree that the fundamental goal of third-person mindreading is to duplicate, reproduce or mimick the mental life of another (the third feature of the idea of mental simulation). Before criticizing the anti-introspectionist approach to third-person mindreading, let me say what is wrong with the third feature of the idea of mental simulation as applied to third-person mindreading.

The goal of mindreading is to form an accurate picture of what our conspecifics think and have in mind. As theory-theorists have, I think, rightly emphasized, what other people *do* might be observable. At least, their bodily movements are. But what they have in mind – their intentions, beliefs, desires, feelings and emotions – is not directly observable in their actions or behavior. Whether or not the mental simulation (pretence) heuristic is a fundamental part of our method for accomplishing third-person mindreading, our success in the task does presuppose a certain amount of psychological similarity between members of the human species. As simulation theorists have emphasized, existing mechanisms of mimicry may serve a useful purpose: they may enhance the probability of success in tasks of understanding the minds of one's conspecifics. For reasons that will emerge soon, neither mimircy nor even imitation, however, can be the process of third-person mindreading.

Why cannot mindreading be mental imitation, let alone mental mimicry? When I try to understand one of your mental states, e.g., one of your beliefs, I try to capture two features of your mental state: its *content* and the *attitude* under which you entertain the content. The only way I can capture the attitude is by applying a mental concept (i.e., the concept of *belief*). The only way I can capture the content is by exploiting a different representation whose content will be suitably related to the content of your belief. Neither the attitude nor the content of a mental state can be copied.

Genes can be copied because, unless they mutate, they replicate. Utterances of sentences of natural and other languages can be copied or reproduced in virtue of their phonological non-semantic properties. Arguably, direct quotation and photocopying are methods for reproducing

utterances. But when it comes to an appreciation of the content or semantic properties of utterances, reproduction or copying is not an option anymore. Only *interpretation* is available. Interpretation can occur across languages or within one language. Unlike direct quotation, indirect quotation is not a mere reproduction. Nor is, of course, translation, which consists in finding one sentence \underline{s} in language L that is synonymous (which has the same meaning) with a distinct sentence s' from language L'. Offering a summary or an exegesis of an utterance, which may take place within one language, involve finding other sentences in one and the same language to help determine the content of the initial utterance. The contents of two or more representations can stand in a *similarity* (or resemblance) relation of which the identity relation is, at best, an extreme limiting case. L

What is true of the contents of linguistic representations is true *mutatis mutandis* of the contents of mental representations. Unlike linguistic representations, mental representations lack phonological properties. Arguably though, assuming, as I do, that mental representations have neurological properties, techniques of brain imaging, might, unlike ordinary mindreading, allow one to reproduce a mental representation in virtue of copying some of its neurological properties. However, this very process in and of itself does not display the meaning or content of the copied mental representation. Only by *interpretation*, not by copying, can one get at the content or meaning of any representation. Only by creatively providing another representation can one interpret the content of a representation.

According to Gordon's (1986, 1995, 1996) anti-introspectionist version of the simulation theory, to ascribe a belief to someone is to engage in *pretence*.¹³ To ascribe a belief to someone else is to mimick the other's mental state by going into a pretend-state. As Gordon (1986, p. 68) puts it, "to attribute a belief to another person is to make an assertion, to state something as a fact, *within the context of practical simulation*. Acquisition of the capacity to attribute beliefs is acquisition of the capacity to make assertions in such a context." I will now argue that the simulation (or pretence) heuristic can indeed contribute to, but cannot constitute a full account of, third person mindreading. It cannot because there is a gap between engaging in pretence and ascribing beliefs.

¹² Cf. Jacob (1987).

¹³ Gordon (1995, p. 53) explicitly links third-person mindreading to the task of actors.

Gordon's (1986, 1995) account originates from Quine's (1960) famous observation that indirect quotation and the attribution of propositional attitudes are "essentially dramatic acts" in which we "cast ourselves in unreal roles" and "we project ourselves into what, from his remark and other indications, we imagine the speaker's state of mind to have been, and then we say what, in our language, is natural and relevant for us in the state thus feigned". Building on Quine's insight, Stich (1983, pp. 84–88) argues that when I attribute a belief to somebody else (as in "Bush believes that p"), I do *two* things at once: first, I "perform a little skit, doing a bit of play acting" in which the "that"-clause "constitutes the script for the play". Secondly, to say that Bush believes that p is to say that he is in a state similar to the state which would play the same typical role if my utterance of sentence "p" had had a typical causal history.

In other words, to say that Bush believes that p is to say that he is in state similar to the state I would be in were I to utter "p". The point is that Stich does not present the first part of his account – the play acting or pretence – as a *full* account of belief-ascription. On Gordon's account, however, in ascribing a belief, I merely perform a simulation, which gives rise to a pretend thought or a pretend assertion. Stich's account involves two steps, not just one like Gordon's. In other words, Gordon fails to tell us what we are supposed to do with our pretend thought or our pretend assertion (see Fuller, 1995).

On Gordon's pure pretence-theoretic account, in ascribing a belief to another person, I engage in pretence, i.e., I entertain a pretend thought or make a pretend assertion: I merely perform a little skit or a bit of play-acting and nothing else. But I think that this pure pretence-theoretic view of third-person mindreading betrays a confusion between engaging in pretence and using pretence for the purpose of representing a representation (or metarepresenting a representation). To ascribe a belief to a third person is to form a thought (or representation) about the other person's thought (or representation). The pure pretence-theoretic account has things backwards: when I ascribe a belief to a third person, I do not make an assertion or entertain a thought within the context of practical simulation. Rather, if I use simulation or pretence at all, I do so within the context of something more complex, i.e., representing a representation. Thinking about someone else's thoughts does not consist in - it does not amount to – either entertaining a pretend thought or making a pretend assertion.

I may ascribe beliefs to a third person for the purpose of deriving from her information about aspects of the world which I either cannot directly perceive myself or do not understand, for the purpose of evaluating her reliability or rationality or for the purpose of arguing with her. When I do so ascribe a belief to a third person, I exercise my concept of belief. It is constitutive of beliefs that they can be misrepresentations: if a mental representation cannot be false, then it is not a belief. So, possession of the concept of belief crucially allows one to entertain complex thoughts (of the type "Bush believes that p") involving embedded representations one knows (or believes) to be false. I can e.g., form the complex thought expressed by the sentence "George W. Bush believes that witches have magical power" or "Aristotle believed that the Earth is at the center of the universe". When I ascribe such beliefs, my beliefascription is not a pretend thought: it is not about some possible imaginary world or situation. It is about George W. Bush's beliefs or Aristotle's beliefs. Since George W. Bush and Aristotle are actual people, my belief-ascriptions describe small portions of the actual world.

To fully appreciate the gap between engaging in pretence and using pretence for the purpose of metarepresenting a representation, consider a piece of fiction, e.g., Shakespeare's Hamlet. It is one thing to engage in pretence for the purpose of enjoying the play and to entertain pretend thoughts about fictitious characters and think (or say that) Hamlet was a 16th century prince of Denmark who was engaged to Ophelia and killed his uncle Claudius, whom he suspected to have murdered his own father. It is another thing to think (or say that) Hamlet, which is the longest of all the plays written by the English Elizabethan playwright, William Shakespeare, depicts a fictitious character named "Hamlet", who was a 16th century prince of Denmark, who was engaged to Ophelia and who killed his uncle Claudius, whom he suspected to have murdered his own father. In entertaining the former pretend thought, one pretends of a fictitious character that it is not fictitious. One's pretend thought is about some possible imaginary world containing several fictitious imaginary characters. In the latter case, one's thought is about the actual world: it is about an actual play, written by an actual playwright, and which anybody can read, see and enjoy by engaging in pretence. Presumably, human children can engage in pretence and games of make-belief before they can fully deploy the concept of belief – as witness famous false belief tasks. Engaging in pretence yields immersed (practical or engaged) forms of third-person mindreading. But human adults have the ability to ascribe, understand and evaluate (for truth and falsity) others people's beliefs in a detached reflective way. This is what lies outside the scope of the simulation theory of third-person mindreading.¹⁴

6. MENTAL SIMULATION AND FIRST-PERSON MINDREADING

Gordon (1996) has offered an account of first-person mindreading consistent with his anti-introspectionist version of the simulation theory and which he calls the "ascent routine" (and which was anticipated by Evans 1982). To see how the "ascent routine" works, consider a question about myself such as: "Do I believe that George W. Bush is the new president of the US?". On the ascent routine model, in order to answer this question, I switch to a logically simpler question: "Is George W. Bush the new president of the US?" This is a question about George W. Bush and US politics, not about myself. According to the ascent routine, if the response to the latter question is "Yes", then the answer to the former question will be "Yes" too. If the answer to the latter question is "No", then so will be the answer to the former question. It is clear why this simple strategy will not work in the general case: at best, it can provide self-ascriptions of beliefs, not sefl-ascriptions of any other attitude (e.g., hope, fear or desire). If the question had been instead "Do I approve of the fact that George W. Bush is the new president of the US?" or "Do I deplore the fact that George W. Bush is the new president of the US?", clearly by using the ascent routine and considering the logically simpler question "Is George W. Bush the new president of the US?", I could not answer such questions. So the ascent routine fails to have the required scope for first-person mindreading.

I now turn to Goldman's introspectionist view of first-person mindreading. Goldman (1989, 1993, 2000) holds an introspectionist version of the simulation theory of third-person mindreading. On his view, one acquires third-person understanding of the mind of others by running a simulation using one's own resources. Thus, third-person mindreading presupposes first-person mindreading. However, Goldman (2000) thinks that simulation fails as an account of psychological self-knowledge.

Goldman's (2000) reason for thinking that mental simulation fails as an account of self-knowledge is quite simple and, I think, decisive. Let us consider the off-line simulation model of my prediction of somebody else's decision. Suppose I am trying to predict a decision you are about

¹⁴ For further discussion, see Jacob (2002) and Dokic (2002).

to make and I am using the pretence-based simulation routine. On this model, I create in myself some pretend beliefs and desires that are supposed to match your beliefs and desires – those beliefs and desires on which you are basing your decision (which I am trying to predict). I feed these pretend states into my practical reasoning system; I turn it on; and it supplies a pretend-decision. Of course, I do not act on this pretend-decision. Rather, I store this pretend-decision in my belief-box as information about you and use it to predict your decision. So, on this model, using a simulation pretence-theoretic strategy to do third-person mindreading depends on or presupposes the ability to do first-person mindreading.

But now the question naturally arises: how did I detect the output of my decision-making system? How did I determine which decision was selected by my practical reasoning system on the basis of the pretend beliefs and desires which served as inputs? Why could I not use the pretence-theoretic simulation heuristic to respond to this question and treat myself in a third-personal way? The reason why I cannot is the following. Were I to use simulation to identify the output of my own decisionmaking system, I would be involved in an infinite regress and I would never succeed in the task. In order to identify the result of my practical reasoning system, I would have to use a second-order simulation heuristic. This second-order simulation heuristic would generate a secondorder output. Now, if I need a second-order simulation heuristic to determine the output of my first-order simulation, then presumably, I could not identify the second-order output of the second-order simulation unless I used a third-order simulation heuristic, which would generate a third-order output, and so on endlessly. But presumably, I can identify the output of my first-order simulation heuristic. And so, I do not use the simulation heuristic to read my own mind.

Presumably, I can simulate myself from the perspective of someone else. But on the one hand, this involves an embedded use of simulation since I cannot directly simulate myself from someone else's perspective: I can only simulate myself from the *simulated* perspective of someone else, i.e., if I simulate the other's perspective. On the other hand, I can simulate myself from the simulated perspective of someone else – for fun – while firmly believing that the other person's beliefs about me are *wrong*, not right. So the question is: how do I read my own mind? What underlies direct immediate introspective self-knowledge? Goldman has

¹⁵ See Dokic (2002, p. 115).

sketched an answer to this question. According to him, we have what he calls "an inner detection mechanism" attuned to identifying our own mental states. In the past, Goldman has argued that propositional attitudes (beliefs and desires), as well as sensory states, have so-called phenomenal properties, i.e., intrinsic qualitative properties. If they do, then these phenomenal properties would presumably be the targets of the inner detection mechanism.

First of all, I want to point out that Goldman's proposal will not get off the ground unless beliefs and desires do indeed have phenomenal properties. I am skeptical. For example, I doubt very much that there is any phenomenology to believing that 5 is a prime number. I think that beliefs can give rise to emotions and emotions do have intrinsic phenomenological properties. But the idea that beliefs have intrinsic phenomenological properties might arise from confusion between beliefs and belief-caused emotions. Plainly, one and the same proposition held under the same attitude (i.e., belief) can give rise to different emotions in two different persons at the same time in virtue of the differences between their belief systems and their values. Think for example of the proposition that Bill Clinton was not impeached in the fall 1998. Presumably, both Kenneth Starr and Bill Clinton can hold the belief that this proposition is true. However, if they do, this single belief will not prompt the same emotional response in both.

Secondly, Goldman's quasi-perceptual model of first-person mindreading seems open to the objections against the traditional empiricist conception of psychological self-knowledge. First of all, if one's knowledge of psychological facts about oneself derives from some quasiperceptual acquaintance with one's own perceptual experiences, then it is questionable whether – as empiricist philosophers (but not Goldman) have claimed – self-knowledge can still meet the epistemic requirements of genuine a priority, let alone infallibility. Secondly, as many philosophers have noticed, whereas vision, audition, olfaction, touch and proprioception can be used to get information about mind-independent facts (involving one's own body and the bodies of others), no inner sense organ provides information about one's own perceptual experiences, let alone about one's own thoughts. Finally, as Harman (1990), Tye (1992), Shoemaker (1996) and many others have noticed, perceptual experiences are introspectively transparent (in a sense of transparency different from the Cartesian sense): they are diaphanous. What one experiences are mind-independent objects, not one's own experiences. One sees, hears or smells non-mental things by having visual, auditive or

olfactory experiences. But one one does not see, hear or smell one's own visual, auditory or olfactory experiences. In other words, the phenomenology of the introspection of one's own visual experience of a bush of blue lavender just is the phenomenology of one's visual experience of a bush of blue lavender. What it is like to introspect and to be aware of one's own visual experience of a bush of blue lavender is nothing other than what it is like to have the visual experience of a bush of blue lavender. Presumably, if introspection of one's visual experience of anything involved some quasi-perceptual process, then introspective awareness of one's visual experience would have a phenomenology of its own – in addition to and above that of the visual experience itself.

More generally, the quasi-perceptual model of self-knowledge seems open to Shoemaker's (1996) powerful objections. Shoemaker (1996, pp. 204–6), who accepts the distinction between perceiving objects and perceiving facts, lists eight conditions all of which are satisfied by the "object perception model" and three of which are satisfied by what he calls the "broad perceptual model." (1) Sense perception results from an organ of perception. (2) Sense perception involves a distinctive phenomenology. (3) Perceptual awareness of facts is mediated by perceptual awareness of objects. (4) Sense perception of an object leads to identification of the object. (5) Sense perception of an object involves a causal relation between a sense organ and the object's intrinsic properties. (6) Objects of perception can be tracked by attention. (7) Perceptual beliefs are caused by the facts involving the perceived objects. (8) Facts involving perceived objects are independent from the perceptual mechanisms.

If, as I think, Shoemaker's eight conditions do fix the boundaries of sense perception, then Goldman's quasi-perceptual model of introspective self-knowledge is in trouble. Whenever one knows what one experiences or what one believes (whether one's belief is occurrent or dispositional), one is not made aware of one's experience or of one's belief by means of a sensory experience of either one's experience or of one's belief. One becomes aware of either one's experience or one's belief by means of a conceptual higher-order representation — or metarepresentation — of one's experience or of one's belief. In other words, one would not be aware of one's experience or of one's belief unless one had the concept of *experience* and the concept of *belief*. In

¹⁶ See Rosenthal (1993, this volume) for elaboration of the notion of higher-

Although I embrace Shoemaker's (1996) critique of the perceptual model of self-knowledge, I do not think that his criticism forces one to accept his further view (which I do not accept) that there is a conceptual (non causal) constitutive relation between one's first-order mental states (experiences and propositional attitudes) and self-knowledge. What Shoemaker (1996, pp. 225-7) calls "self-blindness" is, on his view, a conceptual impossibility. A creature cannot have experiences and propositional attitudes and not be aware that she has them. In other words, it is constitutive of a creature with beliefs and experiences that she cannot be "self-blind". I think that this claim is too strong. I accept Shoemaker's (1968) claim that psychological self-knowledge is "immune to error of misidentification" in the following sense. Psychological selfknowledge is not infallible, but since it is not perceptual, it does not involve a process of self-identification. Although fallible, it is, nonetheless, immune to the kind of error to which perceptual processes are exposed.

In between the perceptual model of self-knowledge and the claim that self-blindness is conceptually impossible, there is, I think, room for an intermediary view according to which self-knowledge is neither perceptual nor conceptually inevitable. One may certainly fail to achieve (at least full) psychological self-awareness. I agree with Shoemaker that one's higher-order representations (or metarepresentations) of one's experiences or beliefs are conceptual, not perceptual, representations. But this is consistent with denying that there is a constitutive link between one's experience or one's belief and one's higher-order reepresentation of one's experience or of one's belief.

REFERENCES

Baron-Cohen, S. (1995). *Mindblindness*, Cambridge, Mass.: MIT Press.

Bermudez, J. (1998). *The Paradox of Self-Consciousness*. Cambridge, Mass.: MIT Press.

Campbell, J. (1994). Past, Space and Self. Cambridge, Mass.: MIT Press.
Carey, S. & Spelke, E. (1994). Domain specific knowledge and conceptual change.
In Mapping the Mind, Domain Specificity in Cognition and Culture, ed. L.
Hirschfeld & S.A. Gelman. Cambridge: Cambridge University Press, pp. 169–200.

order thought for a different purpose, namely for the purpose of accounting for the difference between conscious and unconscious mental states.

- Currie, G. (1995). Visual imagery as the simulation of vision. *Mind and Language* 10(1/2): 25–44.
- Davidson, D. (1984). First-person authority. In Davidson, D. *Subjective, Intersubjective, Objective*. Oxford: Oxford University Press 2001, pp. 3–14.
- Davidson, D. (1987). Knowing one's own mind. In Davidson, D. *Subjective, Inter-subjective, Objective*. Oxford: Oxford University Press 2001, pp. 15–38.
- Dokic, J. (2002). Reply to Pierre Jacob. In *Simulation and Knowledge of Action*, ed. J. Dokic & J. Proust. Benjamin, pp. 111–117.
- Evans, G. (1982) The Varieties of Reference. Oxford: Oxford University Press.
- Fuller, G. (1995). Simulation and psychological concepts. In *Mental Simulation*, ed. M. Davies & T. Stone. Oxford: Blackwell 1995, pp. 19–32.
- Gallese, V. (2003). The manifold nature of interpersonal relations: the quest for a common mechanism. *Philosophical Transactions of the Royal Society of Lon*don 358: 517–28.
- Gallese, V. & Goldman, A. (1998). Mirror neurons and the simulation theory of mind-reading. *Trends in Cognitive Sciences* 2(12): 493–501.
- Gibson, J.J. (1979). The Ecological Approach to Visual Perception. Boston: Houghton-Miffin.
- Goldman, A. (1989). Interpretation psychologized. In *Folk Psychology*, ed. M. Davies & T. Stone. Oxford: Blackwell 1995, pp. 74–99.
- Goldman, A. (1993). The study of folk psychology. *Behavioral and Brain Sciences* 16: 15–28.
- Goldman, A. (2000). The mentalizing folk. In *Metarepresentations: A Multidisciplinary Perspective*, ed. D. Sperber. Oxford: Oxford University Press, pp. 171–196.
- Gopnik, A. & Meltzoff, A. (1997). Words, Thoughts and Theories. Cambridge, Mass.: MIT Press.
- Gopnik, A. & Wellman, H.M. (1992). Why the child's theory of mind really is a theory. In *Mental Simulation*, ed. M. Davies & T. Stone. Oxford: Blackwell 1995, pp. 232–258.
- Gopnik, A. & Wellman, H.M. (1994). The theory theory. In *Mapping the Mind, Domain Specificity in Cognition and Culture*, ed. L. Hirschfeld & S.A. Gelman. Cambridge: Cambridge University Press, pp. 257–293.
- Gordon, R. (1986). Folk psychology as simulation. In *Folk Psychology*, ed. M. Davies & T. Stone. Oxford: Blackwell 1995, pp. 60–73.
- Gordon, R. (1995). Simulation without introspection or inference from me to you. In *Mental Simulation*, ed. M. Davies & T. Stone. Oxford: Blackwell 1995, pp. 53–67.
- Gordon, R. (1996). 'Radical' simulationism. In *Theories of Theories of Mind*, ed. P. Carruthers & P.K. Smith. Cambridge: Cambridge University Press, pp. 11–21.
- Harman, G. (1990). The intrinsic quality of experience. In *The Nature of Consciousness*, ed. N. Block, O. Flanagan & G. Güzeldere. Cambridge, Mass.: MIT Press 1997, pp. 663–675.
- Hauser, M.D., Chomsky, N. & Fitch, W. T. (2002). The faculty of language, what is it, who has it, and how did it evolve. *Science* 298: 1569–79.
- Jacob, P. (1987). Thoughts and belief-ascriptions. *Mind and Language* 2(4): 301–25.

- Jacob, P. (2002). The scope and limits of mental simulation. In *Simulation and Knowledge of Action*, ed. J. Dokic & J. Proust. Benjamin, pp. 87–109.
- Jacob, P. & Jeannerod, M. (2003). Ways of Seeing, the Scope and Limits of Visual Cognition. Oxford: Oxford University Press.
- Jeannerod, M. (1997). The Cognitive Neuroscience of Action. Oxford: Blackwell.
- Jeannerod, M. (2001). Neural simulation of action: a unifying mechanism for motor cognition. *Neuroimage* 14: S103–S109.
- Jeannerod, M. (2002). The mechanism of self-recognition in humans. *Behavioural Brain Research* 142: 1–15.
- Leslie, A.M. (1994). Pretending and believing: issues in the theory of ToMM. *Cognition* 50: 211–38.
- Leslie, A.M. & Thaiss, L. (1992). Domain specificity in conceptual development: neuropsychological evidence from autism. *Cognition* 43: 225–51.
- Meltzoff, A.N. (1995). Understanding intentions of others: reenactment of intended acts by 18 months-old children. *Developmental Psychology* 31: 838–50.
- Meltzoff, A.N. & Moore, M.K. (1977). Imitation of facial and manual gestures by human neonates. *Science* 198: 75–78.
- Meltzoff, A.N. & Moore, M.K. (1997). Explaining facial imitation. A theoretical model. *Early Development and Parenting* 6: 179–192.
- Perrett, D.I., Rolls, E.T. & Caan, W. (1982). Visual neurones responsive to faces in the monkey temporal cortex. *Experimental Brain Research* 47: 329–42.
- Perrett, D.I., Harries, M.H., Bevan, R., Thomas, S., Benson, P.J., Mistlin, A.J., Chitty, A.J., Hietanen, J.K. & Ortega, J.E. (1989). Frameworks of analysis for the neural representation of animate objects and actions. *Journal of Experimental Biology* 146: 87–113.
- Quine, W.V.O. (1960). Word and Object. Cambridge, Mass.: MIT Press.
- Rizzolatti, G., Camarda, R., Fogassi, L., Gentilucci, M., Luppino, G. & Matelli, M. (1988). Functional organization of area 6 in the macaque monkey. II. Area F5 and the control of distal movements. *Experimental Brain Research* 71: 491– 507.
- Rizzolatti, G., Fadiga, L., Gallese, V. & Fogassi, L. (1995). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research* 3: 131–141.
- Rizzolatti, G., Fogassi, L. & Gallese, V. (2000). Cortical mechanisms subserving object grasping and action recognition: a new view on the cortical motor functions. In *The New Cognitive Neurosciences*, ed. M.S. Gazzaniga. Cambridge, Mass.: MIT Press, pp. 539–552.
- Rosenthal, D.M. (1993). Thinking that one thinks. In *Consciousness: Psychological and Philosophical Essays*, ed. M. Davies & G.W. Humphreys. Oxford: Blackwell, pp. 197–223.
- Rosenthal, D.M. (this volume). Consciousness, Interpretation, and Higher-order thought.
- Shoemaker, S. (1968). Self-knowledge and self-awareness. *The Journal of Philoso-phy* 65(19): 555–67.
- Shoemaker, S. (1996) *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
- Spelke, E.S. (2000). Core knowledge. American Psychologist 55: 1233–1243.

- Stich, S. (1983). From Folk Psychology to Cognitive Science. Cambridge, Mass.: MIT Press.
- Stich, S. & Nichols, S. (1992). Simulation or Tacit Theory? In *Folk Psychology*, ed. M. Davies & T. Stone. Oxford: Blackwell 1995, pp. 123–158.
- Stich, S. & Nichols, S. (1995) Second Thoughts on Simulation. In *Mental Simulation*, ed. M. Davies & T. Stone. Oxford: Blackwell, pp. 87–108.
- Stroud, B. (1989). Understanding human knowledge in general. In *Knowledge, Readings in Contemporary Epistemology*, ed. S. Bernecker & F. Dretske. Oxford: Oxford University Press 2000, pp. 307–323.
- Tomasello, M. (2000). *The Cultural Origins of Human Cognition*. Cambridge, Mass.: Harvard University Press.
- Tye, M. (1992). Visual qualia and visual content. In *The Contents of Experience*, ed. T. Crane. Cambridge: Cambridge University Press, pp. 158–176.

B. Current Interdisciplinary Research: Neuroscience and Psychoanalysis