Susan Hurley and Nick Chater eds., Perspectives on Imitation: From Cognitive Neuroscience to Social Science (Cambridge MA: The MIT Press).

Some Reflections on the TT - ST Debate

Ruth Garrett Millikan; University of Connecticut

Goldman tells us that the "theory theory" and the "simulation theory" are different theories concerning "how ordinary people go about the business of attributing mental states." This phrase is ambiguous in ways that may make a difference, I think, both to the controversy between the theory theorists and the simulation theorists and to the question what imitation might have to do with mind reading.

First, the question might be taken to concern the natural ontology of beliefs about mental states. What kind of structure does a belief about a mental state have? Supposing that having a belief about a mental state requires one to have a concept or thought of that kind of mental state, what sort of thing is a thought of a mental state? Is it just like the thought of any other sort of state, say of the state of being old, or the state of being sick, or the state of being wet? Suppose that the mental state to be thought about is an intentional state, and suppose that intentional states are mental representations. Is a representation of a representation as a representation (not just as a vehicle) just another ordinary representation but that happens to have a representation as its object? Or does it require a completely different sort of act of mind?

Second, the question might be taken to concern the ontogeny of the ability to have beliefs about mental states. What steps are in the normal developmental process

that leads to the capacity to think about mental states? Was there perhaps a certain cultural/historical process that eventuated in modern humans acquiring the ability to think about their mental states, as there was a cultural/historical process that eventuated in modern humans acquiring the ability to think about numbers or, at least, say, about negative numbers? Or does each child acquire the ability to think about mental states all on its own, a strong disposition to this having been built in, perhaps, by natural selection?

Third, the question might be taken to concern the natural epistemology of beliefs about mental states. How do people discover what mental states other people are in, or discover what states they themselves are in? Besides knowing what states other people are in, sometimes we can predict what states they will soon be in. And sometimes we can retrodict what states they must have been in previously to account for the emergence of their current states. Sometimes we can predict what states we ourselves will be in given certain future conditions. And sometimes we know something of the etiology of the states we are currently in or have been in the past.

Now the origin of the "theory theory" was a story about the first of these three matters. It was a story about what is involved in having any sort of thought or concept at all, not a story about, specifically, thoughts or concepts of mental states. What kind of structure does a belief about anything have? The story was that beliefs are mental representations, indeed, originally, mental sentences (Sellars, Quine and so forth), and that they acquire their content, they get to be about what they are about, because of their inference relations to one another plus their connections to perceptual input and, some philosophers thought, also to motor output. The concept of a kind of mental state

was the concept of that state and not another for the same reason any other concept was about whatever it was about --namely, because of its role in inference. This sort of theory of what a thought is was aptly called a "theory theory" of the nature of thought because it was exactly the same as the theory developed in the second two fifths of the 20th century concerning the meanings of theoretical terms in a scientific theory. As was forcefully pointed out by Hemple, Sellars, Quine and others, for example, having a concept and having a theory came to much the same thing on this analysis. Or, putting this differently, the difference between changing your beliefs and changing what your thoughts were about (changing your meanings) became moot, or at any rate highly problematic.

Goldman objects to the theory theory as applied to thoughts of mental states partly on the grounds that if it were true, no account could be given of the second problem mentioned above, which concerns the ontogeny of the ability to have beliefs about mental states. Goldman claims that the child could not discover a set of laws concerning mental states by observing the origins and progression of its own mental states because, on the theory theory, to be able to think about its own mental states it must already know what the psychological laws are that define these states. Notice, however, that on the theory theory of concepts, this sort of problem is perfectly general, having nothing to do with the theory theory of thoughts of mental states in particular. If the laws of a theory define the concepts in the theory, it seems one could not reach any theory by performing simple inductions in order to derive its associated laws. The critique, if valid, would challenge the whole of the most characteristic 20th century theory of what thoughts are.

Sellars and Quine had a way out of this dilemma. They took it that we learn to think by being taught to speak. We are taught connections between sentences by our elders and we internalize them. Ordinary people do not develop theories of their own, but slowly learn traditional methods of thought handed down through the generations. Formation of genuinely new theories was another matter. Many, including Sellars --the the original theory theorist about thoughts of mental states-- were explicit about the use of models and analogy in the development of theories. Sellars thought the original model for the ordinary theory of thought was language, and suggested that the development of this new theory took place originally during the history of ideas, not during evolutionary history. But he thought individual children learned about the existence of mental states by being taught correct sentence connections and input conditions for uttering sentences about mental states by their elders and then internalizing these sentences and connections.

Sellars was also explicit about how a person introspects their own current beliefs and desires, and about how they learn to do this. That is, he also had a theory about part of the third question above: How do people discover what mental states they are in? It is not done (as Goldman suggests the theory theorist must hold) by observing one's behavior, but by catching oneself in the state of being disposed candidly to express a certain thought and then prefacing that expression with "I believe" or "I want" or whatever. The fact that an entity is first discovered merely as a theoretical entity does not preclude that one can later learn how to observe it directly, how to make judgments about it directly from experience without inference. In contrast to this, no simulation theorist has, to my knowledge, developed a clear theory of how one knows what one's

own thoughts are, whether they are really one's own, or merely thoughts that one has simulated in the pretend guise of another. What the theory theorist clearly has on his side, concerning both question two and question three above, is the very clear understanding that merely having a mind is not knowing about minds, nor having mental states the same as knowing one has mental states. The general capacity to think of mental states needs to be explained, including the capacity to know what particular mental states one is currently in.

Now I am not disposed to accept the classical theory theorist's view either of the nature of thoughts generally or of the nature of thoughts about thoughts. Nor, of course, am I disposed to accept their view of how children learn to think about thoughts.. But it seems to me that a critique of the classical theory theory needs to go considerably deeper than does Goldman's current analysis. Most important, to oppose the theory theorist on his original ground, one would need to develop a different theory than the classical 20th century theory theory, either of what it is to think about or have a concept of anything --say of dogs, or of the state of being old, and so forth-- or, alternatively, one would need to explain exactly why it is that mental states cannot be thought about in the same sort of way as any other states, and of how they are thought of instead. I think myself that the theory theory of thoughts and concepts is mistaken quite generally. (For this, see Millikan, On Clear and Confused Ideas, Cambridge University Press 2000)). But, moving closer to Goldman's view, there might also be a reason to suppose that there is something peculiar about concepts, not of mental states in general, but at least of intentional mental states, states that seem to be like inner representations.

Sellars took it that our model for thoughts was words and sentences, but not words and sentences classified by vehicle types, but rather as classified by "roles" -- later terminology would have said by "functional roles" or "inferential roles." He did not explain what it is to have a thought of a role, however. What would it be to think of a thought as being a mental sentence that plays a certain role in inference and as typically stimulated by such and sensory input? Would a full-blooded theory theorist have to say that this would involve having a mental name for the sentence, say "Tobermory," and then believing a whole host of psychological laws about what other mental sentences, say Samantha and Melissa and Xavier, when these are believed, generally produce in their wake belief in the sentence Tobermory, and what mental sentences, when believed along with Tobermory, generally produce in their wake still further sentences, for example, Tobias and Melek and Dildar and so forth? One way or another, it seems to me, the theory theorist would need to concede that our ability to think of the inferential role a mental sentence plays must ride piggyback on our own dispositions to make inferences with just such a mental sentence, not on an entirely independent and prior knowledge of what these dispositions are.

Putting this differently, suppose that thinking that someone holds a mental sentence, p, to be true involves thinking that they are likely to believe whatever p immediately implies. This supposition is definitional of the theory theory of what thought is if we spell out the theory theory assumption that believing that someone holds p true involves having a more or less correct theory of what it actually is to hold p true. Now ask, what is it like to know what p immediately implies? The obvious answer would seem to be that this knowledge must, somehow, rest quite immediately on one's having

oneself a set of inference dispositions with regard to the thought that p --not, in the first instance, a set of beliefs about laws of thought but a set of dispositions to obey laws of thought.

Something like this principle would seem to generalize to any theory of what it is to have a thought of a thought if it is assumed that thoughts are mental representations. In order to believe that John plays the trumpet I certainly don't need to be able to play the trumpet myself. But in order to believe that John believes that it is raining I do need to be able myself, if not to believe that it is raining, at least to entertain the thought that it is raining. Surely thinking of a representation, not just as a vehicle, but as something having a known intentional content, requires that I be able to think of or entertain that intentional content myself.

Returning to simulation, if we suppose that merely thinking about a certain content, or entertaining it, involves harboring a representation of that content which is processed, as it is said, "offline," that is, not connected with dispositions to act, as in the case of "on line" beliefs, and if we refer to this sort of offline processing as "simulation" of belief, then it seems to follow that any mental representation theorist will have to agree that the ability to simulate beliefs one doesn't oneself have must lie behind the ability to attribute beliefs to others.

Notice, however, that it does not follow that one might not also simply remember from experience what kind of conclusions one has usually reached from what kinds of experiences or from what kinds of prior beliefs, or remember having been told what kinds of conclusions another has reached, thus concluding what another may think without currently engaging in simulation. Nor does it follow that the ability to know what

kinds of nonintentional mental states tend to have what kinds of outcomes, or what kinds of situations tend to cause what kinds of nonintentional states, depend on concurrent simulation of these states. Only thoughts of intentional mental states succumb to this argument.

And concerning predictions of future intentional states, I think it is important to recognize that prediction of people's future intentions often works backwards by prediction, first, of their future actions. That is, regularity in actions is what we notice first about people, ourselves, perhaps, included. We know what people generally do in what kinds of situations, or what people of a certain culture or class are likely to do, or what a certain individual is likely to do, all by simple induction. If we then think about these people's intentions, it is likely to be in order to explain the behavior we expect rather than predicting their behavior by first knowing their intentions. Thus we are disconcerted when we find certain people, as we say, "unpredictable," for in many ways most people are quite predictable in many broad ways (though not usually in the details of exactly how they will do this or that.) The idea that we use mind reading primarily to predict behavior seems to me quite mistaken. Mostly we use it only to explain behavior after the fact, or to explain behavior that has been predicted by simple induction from previous behavior patterns.

My suggestion, then, is that thinking about intentional mental states probably requires the capacity, at least, to entertain mental representations offline. But so does imagining of all kinds, and so does hypothetical thinking, and so does considering possibilities and hypothesis and so forth. Do all of these derive somehow from the ability to imitate? That seems doubtful. Does the ability to imitate require the ability to think off

line? Well, that might depend, as Goldman seems to agree, on how you define imitation. In neither case, however, it seems to me, is there evidence of any but the most indirect connections between imitation and mind reading specifically.

I feel compelled to add one more skeptical footnote. I do not know where it has been shown that the phenomenon of mirror neurons, mentioned in Goldman's paper and in many of the talks at the Royaumont Abbey meeting on imitation, needs to be interpreted as any different from the well known phenomenon of efferent copy. To interpret it as a phenomenon of efferent copy, all one needs is to assume that efferent copy can predict perceptions of object-centered or aperspectival happenings as well as perspectival happenings. If so, efferent copy could predict, say, that a hand is seen in an object-centered way grasping a nut, disregarding how the hand will be related to the subject's own body. Another part of efferent copy, connected, say, with the "where" (the dorsal) rather than the "what" (the ventral) system, might predict the relation of the grasping to the subject's body. If that were so, there would be no potential causation between seeing another grasping a nut and grasping it oneself. The causation would only be between intending to grasp a nut and the firing of neurons anticipating the seeing of the grasping of a nut. But it may be that the neurologists have their own careful reasons for thinking that is not the way it goes, and I just have not heard them yet.