Understanding Everyday Mental Concepts: Theory or Simulation?

Louise Röska-Hardy, Nordrhein-Westfalen

In the course of explaining and anticipating thought and action, we characterize both ourselves and others in mental terms, e.g. as '*believing* the bus departs at ten' or '*wanting* to go swimming'. The ability to characterize oneself and others in such terms is central to understanding persons. It underwrites the self-attribution of beliefs, desires, emotions and other conscious, occurrent 'mental states' (MS) and subserves MS-attribution to others. How do ordinary people understand the contents of mental state concepts like belief or desire? Philosophers of mind and cognitive scientists have put forward answers based on theory-theory (TT), modularity theory (MT) and simulation theory (ST). Before examining how these approaches explain the contents of MS-concepts, two widespread assumptions concerning everyday psychological attributions should be discussed.

The first concerns the notion that the same MS-concept is attributed in self-attributions and in other-attributions, i.e. the same mental concept is applied to oneself and to others. On this assumption the concepts expressed by mental predicates are unitary and the mental predicates employed in attributions are not ambiguous between first-person and other person uses. The contention that mental state predicates are univocal and mental state concepts unitary is supported by the observation that first-person and third-person mental attributions are symmetric with respect to truth or falsity.

The second assumption concerns whether there is a principled asymmetry between self- and other-attributions of mental states, a point familiar from the discussions of L. Wittgenstein, P.F. Strawson, E. Tugendhat, G. Evans and D. Davidson. That there is an epistemological asymmetry between self-attribution and other-attribution concerning criteria for attribution, e.g. the observation of behavior and speech and the role of inference, has been argued by many. On the asymmetry assumption the grounds for self- and other-attributions of mental states vary, even though the meanings of the MS-predicates employed appear to be univocal. One need not subscribe to a privileged access thesis in order to maintain that there is a principled asymmetry between self- and other-attributions. The notion that MS-concepts are indexical, i.e. tied to a particular point of view, suffices to induce an asymmetry between MS-attributions to oneself and to others.

In attributing a mental state, an attributor characterizes the modality in the case of sensations and perceptions. In the case of beliefs and other "propositional attitudes" both the MS-type ψ and the content p must be characterized. Plausibly, the phrases '*hoping* it is raining' and '*believing* it is raining' characterize different mental states and express different MS-concepts, because the MS-types are different. Whereas, '*believing* it is raining' and '*believing* the sun is shining' are different states, because their contents differ, even though the MS-type is the same. Let's set the question of content to one side and concentrate on MS-concepts. If the attribution of a mental state, whether to oneself or another, consists in characterizing a person's mental condition as being of one type or another, then MS-attribution involves a judgement. Hence, the notion of a MS-type like believing, wanting, etc. is the notion of a mental concept. How do ordinary attributors understand mental concepts like belief or desire?

According to TT, the concepts employed in MS-attribution are theoretical concepts, postulated in explaining and predicting thought and action. In attributing mental states to ourselves and others, we inferentially apply a folk-psychological theory of mind. Depending on the particular version, the theory involved is conceived as a set of generalizations or laws for the deployment of mental concepts (Lewis 1972), or as a theory like any other scientific theory (Churchland 1988; Gopnik 1993). One attributes mental states by inferring their occurrence from observation of behavior and environmental events, i.e. by recognizing their causal-explanatory role in accordance with the theoretical generalizations comprising the theory. Versions of TT differ as to whether the theory in question is acquired through a process of theory formation analogous to scientific theorizing or through learning. However, both versions require that ordinary attributors grasp a mental concept as it is defined by the generalizations of the theory in order to make corresponding MS-attributions.

On TT understanding a mental concept involves mastering theoretical generalizations about causal-inferential relations and MS-representations. On one influential version, mental states like beliefs and desires are defined functionally in terms of their (causal) relations to events in the environment, other mental states and overt behavior. The concept of a mental state is the concept of a particular functional role. Specifying the content of a MS-type on this account thus involves generalizations that make reference to dispositions, causal interactions or subjunctive considerations, as Alvin Goldman has pointed out. These generalizations -- specifically, the causal-functional role the terms play in the generalizations of the folk-psychological theory -- determine the contents of mental concepts (Lewis 1972; Churchland 1988). Consequently, one must master the "theoretical" concepts of the theory in order to employ MS-predicates. Even though the grasp of mental concepts may be "tacit" or "implicit", it is incumbent on the TT to explain how attributors can acquire facility with such concepts. In particular, TT must explain how ordinary attributors acquire and deploy such concepts, for the TT claims that knowledge of the folk-psychological theory of mind grounds MS-attributions.

An important consequence of TT is that mental concepts are defined from what is essentially a third-person or observer point of view. The contents of mental concepts are specified in terms of logical and epistemological relations between external stimuli, mental states and behavior, i.e. on the basis of inference and observation of external stimuli and behavior. They are not essentially linked to a first-person point of view. In this respect mental concepts as defined by TT do not differ in kind from non-mental concepts in theories about the natural world. Clearly, the third-person approach of TT to mental attribution accounts for the assumed uniformity in applying mental concepts. The same theory is used both in self- and other-attributions. By the same token, the TT entails that there is no principled asymmetry between MS-attributions to oneself and to others; for we attribute mental

states both to ourselves and others by means of the folk-psychological theory and theory-mediated inference. However, the question of how one learns to apply folk platitudes to oneself or how theory formation gets started, e.g. where MS-concepts come from, remains open.

Although modularity theories are often construed as versions of TT, even by their proponents, there are good reasons for treating them separately. For they construe cognitive structures like the apparatus of MS-attribution as the result of innate modules, not as the consequence of an acquired theory. On MT mentalizing abilities are not developed from evidence in the course of development, but created from pre-determined representations of input. Leading proponents of modularity, Noam Chomsky, Jerry Fodor and Alan Leslie, differ significantly in their specific positions and respective claims about innate modules. However, the central idea is that the contents of everyday mental concepts in the Theory of Mind (ToM) are part of a special purpose body of knowledge in a mental module, which is innate and matures through a process of ontogenetic development (Fodor 1992; Leslie 1987). Although MT, like TT, construes MS-concepts as abstract theoretical postulates embedded in causal laws, it claims that such concepts are part of our innate endowment as humans. Consequently, their development is genetically determined and not based on theory construction. On MT mental concepts are present from the beginning and merely triggered by experience and maturation. Leslie (1994) postulates several different modules that come on line sequentially in children's developing Theory of Mind. He hypothesizes that a domain-specific mechanism or information processing device, the ToM Module, computes data structures, metarepresentations, which specify attitude, agent, an anchor (an aspect of the real situation) and a pretend state. Modularity theories like Leslie's ToMM hypothesis account for the assumption of uniformity by citing our innate endowment as members of the human species, but offer no further elaboration of mental state concepts. Thus, beyond MT's claim that mental concepts are innate, the little light is shed on ordinary understanding of mental concepts.

Simulation approaches deny that MS-attributions rely on theoretical knowledge and inference. They take the first-person point of view to be essential to mental states and concepts despite different views on the nature of simulation, e.g. on whether it involves analogical inference (Goldman), "ascent routines" (Gordon) or "co-cognition" and rationality assumptions (Heal). Simulationists also differ over whether prior possession of MS-concepts is required (Goldman; Heal) or not (Gordon). The central idea is that one uses one's own cognitive resources to attribute mental states by pretending or imagining oneself to be in the other's position and then generating the thoughts or actions attributed to the other within a simulation.

On Alvin Goldman's introspection-based ST, one attributes mental states to others by using one's own cognitive and inferential mechanisms to match those of another person (1993, 2000). Simulation terminates in an analogical inference from oneself to the other. The end product is a judgement, which classifies an (occurrent) mental state as a token of a particular MS-type and thus requires mental concepts. Goldman proposes an account of mental concepts in terms of introspection or self-monitoring, not simulation. He suggests we have a first-person, introspective understanding of MS-concepts which rests on direct, non-inferential access to our conscious mental states, i.e. inner sense or higher-order perception. Goldman hypothesizes that some sort of intrinsic (non-relational) and categorial (non-dispositional) properties of mental states are available to the cognitive system. These internally detectable properties of MS-tokens are definitive of a particular MS-type and form the basis for our understanding of mental state concepts.

The asymmetry of self- and other-attributions is a central tenet of Goldman's account, for the contents of mental concepts are anchored in "phenomenologically disinctive" MS-properties which are introspectively detectable. On this view the question of the uniformity of MS-attributions stands and falls with the viability of the analogical inference from oneself to the other. Specifically, what reason is there on Goldman's view to assume that others have mental states like one's own? This is the conceptual problem of other minds. In response Goldman has proposed a dual-representation hypothesis for MS-concepts like desires, suggesting that one develops two sorts of MS-representations: representations of inner features and representations of behavioral characteristics, perhaps based on resonance phenomena like "mirror neurons" (2000). However, this response still does not meet the "other minds"-objection which undermines any uniformity assumption.

Robert Gordon takes the simulation approach itself to have direct implications for our understanding of mental concepts. He denies that simulation relies on introspective access to one's own mental states or that it involves an implicit analogical inference (1995). He introduces the notion of an 'ascent routine' to account for self- and other-attributions of mental states without introspection or analogical inference. An ascent routine is a procedure which allows one to get the answer to a question about one's mental condition by answering a question that is not about oneself or mental states, e.g. "Do you *believe* that Mickey Mouse has a tail?", by asking oneself an "outward-looking" question about the world, "*Does* Mickey Mouse have a tail?".

According to Gordon, an attributor "recenters her cognitive map" on the other so that the first-person pronoun 'I' refers to the individual on whom the attributor's egocentric map has been recentered. After this imaginative transformation into the other, a simulator *directly* attributes the belief or decision generated within the scope of simulation to the other *via* an ascent routine, thus obviating the need for an analogical inference. Other-attribution is, in essence, a case of mental "self"-attribution to oneself-as-the-other within the context of simulation. This is because embedding an ascent routine within a simulation allows one to attribute mental states to the other „directly". On Gordon's account an egocentric shift on the part of the attributor lies at the core of simulation and the ascent routine is the key to other-attribution *and* self-attribution. Note that Gordon's ascent routine ST accommodates the uniformity assumption by default, since self- attributions *and* other-attributions are cases of „self"-attribution *via* ascent routines.

Gordon suggests that ascent routines provide the basis for an account of MS-concepts in terms of simulation and indicate how we go about mastering mental state concepts, because they provide a way of reconceptualizing pains and beliefs as having a mental location (1996). However, this is not the case. The meanings of mental predicates and a grasp of mental concepts must be presupposed in order to get an ascent routine started. One must understand that the initial question in the example is about belief, not hope or desire. Thus, the ascent routine procedure does not supply the foundation for an account of mental state concepts. It does accommodate the indexicality intuition through the notion of a mental location, but it falls short of capturing the asymmetry of mental attribution. For mental attribution is essentially self-attribution either directly *via* ascent routine or to oneself-

as-the-other within a simulation. It remains unclear how genuine other-attribution is to be understood. The crucial assumption that one can mentally "transform oneself into the other", "recenter one's cognitive map" or "become the other" remains a metaphor, leaving any asymmetry between self- and other-attribution of mental states a mystery. In the end, imaginative transformation *cum* ascent routine appears to presuppose an account of the contents of mental concepts, rather than providing one.

Neither TT, MT or ST accommodates both the uniformity and the asymmetry assumption. However, it is argued that they are well-motivated and cannot be dismissed out of hand, if we are to explicate how people understand everyday mental states. As things stand, the answer to the question whether our understanding of everyday mental concepts is a matter of theory or simulation is 'neither of the above'.

References

Churchland 1988 *Matter and Consciousness*, Cambridge, MA: MIT Press.

Fodor, J. 1992 „A Theory of the Child's Theory of Mind", *Cognition,* 44, 283-96.

Goldman, A. 1993 „The Psychology of Folk Psychology", *Behavioral and Brain Sciences,* 16, 15-28.

Goldman, A. 2000 „Folk Psychology and Mental Concepts", *Protosociology,* 14, 4-25.

Gopnik, A. 1993 „How We Know Our Own Minds: the Illusion of First-Person Knowledge of Intentionality", *Behavioral and Brain Sciences,* 16, 1-14.

Gordon, R. 1995 „Simulation without Introspection or Inference from Me to You", in *Mental Simulation*, M. Davies and T. Stone (eds.), Oxford: Blackwell's.

Gordon, R. 1996 „'Radical' Simulationism", in: *Theories of Theories of Mind*, P. Carruthers and P. K. Smith (eds.), Cambridge: Cambridge University Press.

Heal, J. 1998 „Co-cognition and Off-line Simulation: Two Ways of Understanding the Simulation Approach", *Mind and Language,* 13, 477-498.

Leslie, A. 1987 „Pretense and Recognition: the 'Origins' of the Theory of Mind", *Psychological Review* 94, 412-26.

Leslie, A. 1994 "Pretending and Believing: Issues in the Theory of ToMM", *Cognition,* 50, 211-38.

Lewis, D. 1972 "Psychophysical and Theoretical Identifications", *Australasian Journal of Philosophy,* 50, 249-58.

Strawson, P. F. 1959 *Individuals*, London: Methuen.