Simulation Theory With or Without Introspection: An Expressivist Compromise

John Michael, Copenhagen, Denmark

JOAL@dpu.dk

In this paper, I will defend Robert Gordon's non-introspectionist version of the simulation theory of social cognition against the criticism put forth by Alvin Goldman, who argues that simulation theory must include an introspectionist account of mental concepts. My strategy will be to isolate the compelling part of Goldman's challenge and then consider conceptual options for meeting it without turning to full-blown introspectionism, i.e. by taking on a deflationary account of introspection, based upon Wittgenstein's expressivist conception of self-ascription. But first I will briefly introduce and contextualize simulation theory.

1. Background: theory theory versus simulation theory

Theoretical debate and empirical research into social cognition and mental concepts have been dominated by two theories: theory theory and simulation theory. According to theory theory, social cognition involves two central components:

i) Mental concepts to denote mental states, conceived as unobservable entities causing behavior.

ii) Nomological generalizations linking mental states among each other and to perception and behavior.

Obviously, two components are closely linked. Obviously, theory theory yields a functionalist account of mental concepts, which highlights criteria that are accessible from the third-person perspective, thus marginalizing introspection.

The basic insight of simulation theory is that we don't need to represent these nomological functional relations since we embody them insofar as we are similar to other people. We can simply put ourselves in another person's shoes and see how we would act, what we would think, or how we would feel, and then expect the same of them. Various versions of simulation theory differ however with respect to the account of mental concepts they derive from this basic picture.

Before I explain the difference, I will point out that, although the debate is primarily about third-person ascription, first-person ascription must also be part of the picture of mental concepts. Specifically, there should in fact be a symmetry or stability between the contents of mental concepts used in first- and third-person ascriptions, otherwise we would not understand other people when they talk tell us what they think or feel. Both theories therefore include accounts of first-person ascription. For simulation theory, this is all the more important since it makes first-person ascription primary. After all, the idea is that first-person embodiment of the psychological apparatus constituting nomological relations obviates the need for representations thereof. The question at stake, then, is what kind of access people need to have to these embodied relations in order to exploit them in simulations for social cognition. Specifically, does it make sense to regard this kind of access as introspective?

2. Simulation with or without introspection?

Alvin Goldman (2006) rejects only the second component of theory theory, namely the representations of nomological relations supposedly used to derive predictions about behavior from constellations of mental states. But he thinks that mental concepts need to be used (a) in order to set up a simulation – i.e. in order to ascribe a constellation of beliefs and desires to someone so that we can simulate their perspective and see how we would act – and also (b) to exploit a simulation – i.e. to identify the state that is the output of the simulation in order to ascribe it to the target person. But that means that he needs an account of mental concepts that does not rest on representations of nomological relations. He therefore turns to internally accessible criteria, i.e. introspectionism. I will come back to his proposal below.

Robert Gordon's (1995) "radical simulation" theory rejects both components of theory theory, i.e. Gordon denies that mental concepts play a role in setting up or exploiting a simulation. Gordon appeals to Gareth Evans' ascent routine: if you are in a position to assert "p", you are in a position to assert, "I believe that p". In other words, you can tack on "I believe that …" reliably without even possessing a full concept of belief, and the result will function as a belief would. According to Gordon, you can assert "p" in a simulation of someone else and tack on a sort of tag to the effect that p is their assertion and not yours, and it will function reliably like an ascription of a belief. Many people find this attractive because they, like Gordon, find it phenomenologically plausible that we think about situations rather than about people's minds when we are trying to understand them. Unfortunately, there are some problems with Gordon's view. I will focus on just one of them, which is the basis of Goldman's case for introspectionism.

Evans' account is designed for beliefs; it accomplishes a transition from a first-order utterance about the world to a second-order utterance about a belief about the world. But can we use an ascent routine to self-ascribe other propositional attitudes? If someone asks me "do you hope that p?" there is no obvious way to apply the ascent routine model to answer the question. This criticism indeed points to a limitation, namely that although the ascent routine is a good explanation of how the content of propositional attitudes can be redeployed in an ascription, it does not explain how the attitudes themselves can be identified for the purposes of ascription.

To see this, consider a simple example of the kind of social cognitive achievement that Gordon and Goldman both want to explain. Sammy utters the sentence "The Yankees will win." Gertrude understands this to be an expression of despair rather than a mere prediction or an expression of hope. Gertrude then accordingly draws inferences about how Sammy will act in a range of situations, e.g. being invited to a glass of champagne, hearing the

surprising news that the Yankees have lost, seeing a person wearing a Yankees shirt, etc.

The distinctions among different mental concepts – such as BELIEF, DESPAIR, and HOPE – enable one to draw these inferences. Gordon wants to explain that via ascent routines. The problem is that insofar as the content of the mental states in question is the same (hoping/despairing/predicting that the Yankees will win), mere redeployment of the content will not help Gertrude to draw the different inferences. She must take into consideration some other properties of the state being interpreted or ascribed, such as the attitude toward that content and/or the intensity of the attitude. Goldman's point is that in order to take such properties into consideration, Gertrude must become aware of them in some way, and Gordon's ascent-model gives no help with that.

3. Towards an expressivist account of introspection

Goldman wants to avoid positing that there is an internally identifiable marker for each propositional attitude, since he thinks that would lead to an unparsimonious explosion of internal markers. Instead, he makes the plausible assumption that we are sensitive to a finite set of internally accessible criteria or parameters, and that different propositional attitudes are constituted by different combinations of settings of these parameters. Goldman's tentative proposal envisages just three such parameters, namely a doxastic, a valence and a bodily feeling parameter. HOPE, for example, would be constituted by a relatively positive setting on the valence parameters plus a relatively uncertain setting on the doxastic parameter, and perhaps some proprioceptively accessible typical bodily changes (e.g. increased heart-rate, upright posture). DESPAIR, in contrast, would combine a negative setting of the valence parameter with near-certainty on the doxastic parameter, and typical bodily changes etc. I think that this is a reasonable tentative proposal concerning what internal parameters we might be sensitive to. But it is not the case that introspection – in a full-blown sense is the only way to access these parameters. On the contrary, I will argue that there are conceptual resources for articulating how we monitor and self-ascribe these parameters of our mental states without detecting them or directly becoming aware of them, and that the relevant empirical evidence favors such a deflationary proposal.

First of all let me point out that the ascent routine model and thus Gordon's version of simulation theory, is a species of *expressivism*, which is an essentially Wittgensteinian view of self-ascription. It is based upon self-ascription via self-expression, e.g. I scream "that hurts" when you stick a knife into my leg or I cautiously say "I believe Stockholm is larger than Oslo" to express uncertainty. In both cases, the self-ascription flows from the internal state in the same way that a facial expression flows from an internal state, i.e. without any attempt to detect that state.

This account may qualify as introspectionist insofar as it postulates a special, first-person mode of access to one's mental states, but it is deflationary insofar as it does not conceive of this access as perceptual. It may sound like a mere trick that cannot serve as a model of how we usually keep track of our mental states. But there is in fact relevant empirical evidence suggesting that there are states filling the functional roles postulated by Goldman's three introspectively accessible parameters and which can influence our decision-making, expectations and inferential processes via expression rather than via full-blown introspection, i.e. without our detecting them or directly becoming aware of them.

I will focus on the doxastic parameter here, and will say just a bit about the other two towards the end. The kind of mental state or process that would be suited to play the role of a doxastic parameter would have to produce an assessment of our state of confidence in the quality of information about the world being used in decision-making, expectation-formation or inferential processes. There is in fact lots of empirical work on such "epistemic feelings" that help us in judging the adequacy of a particular response or evaluating the ease or difficulty of learning some new information or of recalling some previously learned information (Proust 2006). Epistemic feelings provide us with internal cues that lead us to produce different behaviors. An uncertainty cue, for example, would lead one to seek more information, to hesitate, to repeat a learning process or to modify it, not make a high wager on one's guess. In other words, they are internal cues that enable us to distinguish among degrees of certainty.

The presence of such skills in species that lack theory of mind abilities speaks against the idea that their use depends upon their being linked up with mental concepts (Proust 2006). So far, this supports Goldman's general picture – since he wants the doxastic parameters to be a *component* of the mental states picked out by propositional attitude concepts, he obviously needs to exclude the possibility that they *presuppose* mental concepts. But, crucially, it is not at all clear that the exploitation of these cues requires us to be aware of them. They may influence our behavior and/or thought processes without our being aware of them, and we may simply be aware of the behavior and/or thought processes that they dispose us to without our being aware of them. If this is the case, then, *a fortiori*, the same would be true of self-ascriptions occurring within the contexts of simulating other people. Then Goldman's analysis of the components of propositional attitudes would be right, but he would be wrong in claiming that we must introspectively access them in order for them to influence our ascriptions.

There are some studies that are relevant to this issue. Persaud et al. (2007) reported a series of studies in which participants performed various tasks under uncertainty, such as pack-selection in the Iowa Gambling Task and visual discrimination in blindsight, and then placed wagers upon their decisions. These tasks are interesting because the participants say they are making blind guesses but perform significantly better than chance, revealing that they have unconscious hunches that are influencing their decisions. Crucially, the participants did not maximize their winnings by placing higher wagers when they had made correct guesses, thus suggesting that they could not distinguish between decisions based upon hunches and decisions that were blind guesses – they were not only unaware of such a distinction, but their wagering decisions also failed to reflect any unconscious tracking of such a distinction. This favors Goldman, since it suggests that the epistemic feelings do not make any difference upon wagering in the absence of awareness.

There was however a condition in which participants in an Iowa Gambling Task – before placing their wagers – ranked the payoff likelihood of packs from -10 to +10 and stated which packs they would prefer to pick from if they had to choose just one. In this condition, their wagers tracked their performance much better, thus maximizing their winnings. This suggests that answering these questions caused them to become aware of their epistemic feelings toward the packs. But – and here is the decisive

point – the way in which the participants in the Persaud study become aware of their epistemic feelings about the packs was by ranking the packs and predicting their own behavior, not by directly looking into their own minds. It is reasonable to interpret this as an indirect access to their epistemic feelings via the behavior that would express those epistemic feelings.

Finally, I will say just a bit concerning the valence and bodily feeling parameters. I take it to be a clear-cut case that preferences and aversions (i.e. states functioning as a valence parameter), and multifarious bodily states (i.e. the bodily state parameters) can influence our thought processes and behavior without our being aware of it. Moreover, they do so in a predictable way, such that we can reliably anticipate how we will think, feel, decide and act when we are in those states – regardless of whether we are aware that they are the cause. In the context of a simulation of someone else, this kind of indirect access to emotional and other bodily states via their expressions enables us to anticipate their actions. All that this proposal requires is that we mirror others' emotions and various bodily states, and that this mirroring causally contribute to our understanding others actions, thoughts and emotions. And indeed there is lots of evidence that thinking about various kinds of perceptual, motor or affective experiences, or observing them in others, causes our own perceptual, motor or affective systems to be activated as they would if we were performing the movements or having the sensory or affective experiences ourselves. And this resonance or embodiment seems to be necessary for understanding others' intentions or emotions and to influence various kinds of conceptual processing (Bastiaansen et al. 2009).

Literature

Bastiaansen, J., M. Thioux and C. Keysers 2009 "Evidence for mirror systems in emotions." *Philosophical Transactions of the Royal Society* B 364: 2391-2404.

Goldman, Alvin 2006 *Simulating Minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.

Persaud, Navindra, Peter Mcleod, Alan Cowey 2007 "Post-decision wagering objectively measures awareness." Nature Neuroscience 10(2): 257-261.

Proust, Joelle 2006 "Rationality and metacognition in non-human animals", in: Hurley, Susan, and Matthew Nudds, (eds.) *Rational Animals*. Oxford University Press: Oxford.

Gordon, R. (1995). Simulation without introspection or inference from me to you. In Stone, T.; Davies, M. (Eds.). *Mental Simulation: Evaluations and Applications*. Oxford: Blackwell.