

Simulation, Theory and Collapse

Bill Wringe

Received: 10 March 2008 / Accepted: 20 August 2008 / Published online: 27 September 2008
© Springer Science+Business Media B.V. 2008

Abstract Recent philosophical discussions of our capacity to attribute mental states to other human beings, and to produce accurate predictions and informative explanations of their behavior which make reference to the content of those states have focused on two apparently contrasting ways in which we might hope to account for these abilities. The first is that of regarding our competence as being under-girded by our grasp of a tacit psychological theory. The second builds on the idea that in trying to get a grip on the mental lives of others we might be able to draw on the fact that we are ourselves subjects of mental states in order to simulate their mental processes. Call these the *theory view* and the *simulation view*. In this paper I wish to discuss an argument—which I shall call Collapse—to the effect that *if* our capacities can be explained in the way that the simulationist supposes then they can also be explained along lines that the advocate of the theory view favours. I am not the first person with simulationist sympathies to have addressed this argument. However, my response is somewhat less concessive than others in the literature: while they attempt to soften its force by attempting to reformulate the simulationist view in a way that evades the conclusion of the argument, I attempt to meet it head on and to show that it does not even succeed in refuting the version of simulationism which it takes as its target.

1 Introduction: Simulationism and (One of) its Critics

Recent philosophical discussions of our capacity to attribute mental states to other human beings, and to produce accurate predictions and informative explanations of their behavior which make reference to the content of those states have focused on two apparently contrasting ways in which we might hope to account for these

B. Wringe (✉)

Department of Philosophy, Bilkent University, 06800 Bilkent, Ankara, Turkey
e-mail: wringe@bilkent.edu.tr

abilities.¹ The first follows a strategy which is standard in many parts of cognitive science: that of regarding our competence as being under-girded by our grasp of a tacit psychological theory. The second builds on the idea that in trying to get a grip on the mental lives of others we might be able to draw on the fact that we are ourselves subjects of mental states in order to simulate their mental processes. Using a now familiar shorthand I shall refer to these two approaches as the *theory view* and the *simulation view*.

In this paper I wish to discuss an argument—which I shall call Collapse—to the effect that *if* our capacities can be explained in the way that the simulationist supposes then they can also be explained along lines that the advocate of the theory view favours. I am not the first person with simulationist sympathies to have addressed this argument. However, my response is somewhat less concessive than others in the literature: while they attempt to soften its force by attempting to reformulate the simulationist view in a way that evades the conclusion of the argument, I attempt to meet it head on and to show that it does not even succeed in refuting the version of simulationism which it takes as its target.²

Many of those who have been involved in the theory view/simulation view debate have moved to a position on which they accept that some combination of theory and simulation is required to fully explain the capacities with which we are concerned here.³ I shall call views of this sort *mixed views*. Given the possibility and current popularity of mixed views one might think that a focus on the Collapse argument was unhelpful, since it seems that this argument takes as its target only a position at the extreme end of a spectrum of possible views. I disagree. As I shall try to show in the next section, if Collapse is a good argument it undercuts mixed views, as well as purely simulation based views. It certainly does not provide—as some might be tempted to think—a line of argument in support of a mixed view.⁴

2 Some Initial Distinctions: Theory and Simulation, Collapse and Mixed Views

My characterisations of the theory view, the simulation view, and mixed views have been very schematic. In the interest of making it clear how my characterisations relate to existing views, I shall expand on these characterizations a little. Doing so will also help me to make good my claim about the relation between Collapse and mixed views.

First, then, the theory view. I take it to be constitutive of a commitment to the theory view that one takes the tacit theory which underlies our capacity to ascribe mental states and to produce accurate predictions and informative mentalistic

¹ Some early discussions of these topics focus entirely on our capacity to predict and explain the behaviour of others. However, following Morton (1996), I take this focus to be unduly narrow.

² For ‘Collapse’ see Davies (1994), Davies and Stone (2001) and for responses Heal (1994) and Ravenscroft (2003).

³ See for example, Heal (1996a), Carruthers (1996), Nicholls and Stich (2003). I talk of ‘mixed views’ rather than ‘the mixed view’ here, since there is clearly a spectrum of different possibilities.

⁴ As was suggested to me in conversation by Mark Ashton Smith.

predictions to be (as I have already said) a *psychological* theory. To count as a psychological in the sense in which I am understanding this term, the theory must make ineliminable and non-trivial use of *representational concepts* (Rather than discuss at length what it takes for a state to count as representational, I shall allow any theory whose explicit statement requires reference to one or more representational states concepts to count as a psychological theory. Theories which refer to beliefs and desires are the most obvious example, but need not be the only ones.).

This is intended to be a fairly liberal characterization of the theory view. However, it is not so liberal that *any* account of mental state ascription automatically counts as a version of the theory view.⁵ Consider, for example, Heal's co-cognition based theory of mental state attribution.⁶ On Heal's view, my ability to ascribe to you thoughts about cupcakes involves my deployment of the same conceptual capacities as you use in thinking about cupcakes. Arguably, the conceptual capacities that you draw on in thinking those thoughts may involve a theory about cupcakes. If so, then in order to ascribe to you thoughts about cupcakes, I need a theory. But this does not make Heal's co-cognition account a version of the theory view, because a theory about cupcakes need not be a theory which involves concepts of representational states, and hence not a psychological theory.⁷

Secondly, mixed views and their relationship to the Collapse argument. As Nicholls and Stich have pointed out, one reaction to debates between advocates of the theory view and the simulation view is to think that debates about our mentalising capacities are obscured rather than made clearer by insisting on a clear cut distinction between theory-based and simulation-based views.⁸ We should see these as extreme ends of a spectrum of possible accounts of how we ascribe mental states. At one end of the spectrum there are information-rich—theory-based—accounts; at the other end there are information-poor—simulation-based accounts, and then there are a range of intermediate accounts which involve some theory and some simulation. Such views are mixed views.

The line of thought I have just described seems to have something in common with Collapse. Both suggest that the theory view/simulation view dichotomy gets in the way of a clear view of what mentalising requires. But we should not conflate Collapse with any version of the mixed view. There are two reasons for this. First, we should distinguish between views and arguments in favour of views. Collapse is an argument; and one might even take it to be a valid argument even if one thought—along with some Wittgensteinians—that neither simulation nor theory plays much role in mental state ascription.⁹

Second, and more importantly: if a mixed view is to be taken to be interestingly different from the theory view, then an advocate of that view must hold that there are some aspects of our mentalising capacity that are not correctly explained by

⁵ At least, in the absence of a successful argument along the lines of Collapse.

⁶ Heal (1998b); see also Wringe (2003).

⁷ Thanks to an anonymous referee for Erkenntnis for making me clarify this.

⁸ Nicholls and Stich (2003, pp. 131–132).

⁹ See for example Hutto (2004).

reference to our possession of a theory. By contrast, if Collapse is correct, advocates of the mixed view make too big a concession to simulationism: any aspect of our mentalising which can be accounted for by reference to our possession of a capacity to simulate can *ipso facto* be accounted for by reference to our possession of a theory.¹⁰ In other words, Collapse undercuts mixed views rather than supporting them.¹¹

This point is worth insisting on, since it is sometimes thought that the impact of Collapse can be discounted if one thinks that some combination of theory and simulation must be involved in our capacity to ascribe mental states. If I am right, those attracted to mixed views need to find some way of drawing the teeth of Collapse.

3 Arguing for ‘Collapse’

I shall now state the Collapse argument as briefly as I can.¹² Collapse relies heavily on an analysis of the notion of tacit knowledge which has been proposed by Martin Davies.¹³ On Davies’ account, an individual has tacit knowledge of a certain theory, and deploys it in performing a particular task just in case, in performing the task in question, he or she goes through a set of states whose causal relationships are isomorphic with the logical relationships that would exist in a logical derivation of a result corresponding to a correct performance of the task from the axioms of a theory. (Davies has in mind here as a paradigm the derivation of the truth conditions of a sentence from the axioms of a semantic theory.)

Collapse proceeds by arguing that an individual who made a mental state attribution via a process of sub-personal level of simulation would, in doing so, instantiate a set of internal states whose causal relations that would mirror the logical relations that would be involved in the derivation of the attribution from a

¹⁰ One slight wrinkle in this account demands attention here. One might take it that ascriptions of mental states involve the possession of mental state concepts, and have a theory of concept possession on which possessing a concept of X involves having a theory about X’s. Simulationists should, I think, resist such theories of concept possession. They can do so even if they are, like Heal (1996b), tempted by and committed to inferentialist theories of concept possession of the sort put forward by Peacocke (1992), provided that they do not allow every set of inferential dispositions to count as a theory. One principled reason for such a refusal might be thinking that it is possible to distinguish between canonical and non-canonical inferences involving a concept. Another might be thinking that there is a *prima facie* case for holding that inferential dispositions involving normative concepts should not be counted as instances of theory possession and taking it that the inferential concepts constitutive of the possession of mental state concepts have this feature.

¹¹ I would like to thank an anonymous referee for *Erkenntnis* for persuading me of the need to be clearer about this.

¹² Thanks again to the same anonymous referee for persuading me that an earlier presentation was *too* brief.

¹³ See, Davies (1987). I shall take for granted that Davies’ account is the best available. The account is certainly widely cited, and as far as I know, no advocates of the simulationist view have advocated a more plausible alternative. If it turns out to be unsatisfactory in some respect, that would in one sense be grist to my mill, insofar as it would further undermine the case against simulationism; but it would make the present paper otiose.

folk-psychological theory. This is because, in simulating, an individual would go through a series of simulated states that would eventually result in them attributing a mental state to another thinker. There would presumably be causal relations between these states: simulationists see simulation as a causal process. There would also be logical relationships among the contents of these states, since it is a key part of the simulationist position that when we simulate we typically simulate rational deliberative processes. Since this is so there would also be logical relationships between a set of statements attributing an initial set of mental states to the simulated individual and a range of subsequent states involved in the derivation. But a statement of these relationships would, in effect, be a derivation of a claim correctly attributing a mental state to the individual being simulated from some initial claims about the contents of his or her thoughts.

In other words, the simulator would be going through a set of causally related states which were isomorphic with a logical derivation of a correct mental state attribution. But if so, then, given Davies' account of tacit knowledge the individual would be deploying tacit knowledge of a psychological theory in making a mental state attribution. Hence the claim that the simulationist account collapses into a version of the theory account.

From a purely dialectical point of view, a good response to Collapse should have three features. First, it should grant the proponent of Collapse both his/her preferred way of understanding the notion of tacit knowledge, and the claim that it is in principle possible to formulate a version of folk psychology which is sufficiently precise to allow for the possibility of there being a logical derivation issuing in an ascription of a folk psychological state and that the precise details of what is involved in simulation at the sub-personal level do not make a great deal of difference to the viability of Collapse. We should also allow that there is some psychologically realistic way in which this theory can be implemented.¹⁴ Secondly, it should impose a minimum of constraints on how the simulationist formulates his/her view. The response to Collapse which I shall be putting forward here is superior to other responses in the literature in these respects.¹⁵

4 Responding to Collapse

A proponent of simulationism should respond to Collapse by focusing carefully on the question of what is really at issue in the theory/simulation debate. That debate is about how certain human capacities are to be *explained*. In particular, we are concerned with finding the *best* explanation of these capacities. Even if we grant the advocate of Collapse everything that she wants, her arguments cannot show that the

¹⁴ Morton (2003) suggests that this may be a problem for the theory view. Thanks to an anonymous referee for Erkenntnis for drawing my attention to this.

¹⁵ Some examples of responses which fail to respect this constraint: Goldman's suggestion that we should discard Davies account of tacit knowledge (Goldman 2006); Heal's resort to uncodifiability (Heal 1998a); Ravenscroft's suggestion that simulation should concentrate on theoretical deliberation rather than practical deliberation (Ravenscroft 2003); and Heal's retreat to personal level simulation (Heal 1994).

theory view provides a better explanation of our capacities for ascribing mental states.

To see why, we need to consider carefully what makes for a good explanation of a human capacity and what makes for good explanation in general. In his context we can set aside truth as a criterion of explanatory goodness. For if Collapse is sound, then it shows that *if* a certain kind of simulationist explanation of mental state attribution is correct, then a theory theorists account is correct as well. In other words, it shows that, given certain assumptions, the two explanations stand on an equal footing as far as truth is concerned.

A further criterion to consider is explanatory depth. Collapse seems to show that, if a certain kind of simulationist story about folk psychological attribution is correct, then a certain kind of theory-theorist account must be as well. The argument seems particularly elegant precisely because it seems to show that if the simulationist account is true, then a theory-theorist account must be true *because* the simulationist account is true; or as one might also put it, the theory theory account is true *in virtue* of the truth of the simulationist account. By contrast, there seems no way of showing that a simulationist account would have to be true *in virtue of* the truth of a theory-theorists account.

However, this apparent strength of the Collapse argument turns out to be its Achilles heel. For, while the notion of explanatory depth is somewhat vague and intuitive, the following seems plausible:

Asymmetry: Of two alternative true explanations, *A* and *B* of a phenomenon *P*, where *A* is true in virtue of the truth of *B*, while *B* is not true in virtue of the truth of *A*, *B* has the greater explanatory depth.

If Asymmetry is correct, then we have good reason to think that the simulationist account of mental state ascription has greater explanatory depth than the theory theorist's account. Asymmetry appeals to the notion of one explanation's being true in virtue of the truth of another explanation. However, the 'in virtue of' locution is not entirely transparent. A good candidate for explicating it itself appeals to the notion of explanation: one explanation is true in virtue of the truth of a second explanation if the first explanation explains not only the phenomenon in question, but also explains why the second explanation explains the phenomenon.

The plausibility of Asymmetry, thus explicated, can be seen by considering a mundane example. I phone a colleague in her office. She is not there. I then notice that her car is not in its usual parking place. I infer that she has not come to the office today since she never travels anywhere except by car. I take this to explain why I am unable to contact her on her office phone. I then phone several other people, all of whom are not in their office. I check the date and realize that today is a national holiday, and infer that no-one is in their office. It seems plausible that I now have a deeper explanation of my colleague's absence. The explanation is deeper because the explanans of my previous explanation (that my colleague is not in her office) is itself explained in this case by the fact that, today being a national holiday, no-one is in their office.

Asymmetry faces a well-known problem. Consider two alternative explanations of why I am unable to force a certain peg into a particular hole. The first explanation

appeals to the electrostatic features of the particles which make up peg and the surroundings of the hole at the atomic level. The second explanation adverts to the fact that the hole is round while the peg is square.¹⁶ If we accept Asymmetry we seem forced to say that the first explanation is deeper than the second. But it also seems plausible that the second explanation is a better explanation.

We should bite the bullet here and accept, as Asymmetry suggests, that the electrostatic explanation is deeper than the geometric one. But we should also say that depth is only one of many explanatory virtues. Another virtue in an explanation is simplicity. Sometimes gains in explanatory simplicity more than compensate for losses in explanatory depth. How far this is true may depend on the purposes for which we want the explanation.

5 Two Ripostes

So far, I have argued that even if Collapse succeeds, the simulationist explanation of our capacity to ascribe folk-psychological states is—if true—deeper, and thus better, than the theory theorist's account. However, I have also conceded that depth is only one of a number of different explanatory virtues, and that it can be outweighed on occasion by other virtues. This leaves the proponent of Collapse the option of responding that the greater explanatory depth of the simulationist account is outweighed by some compensating virtue of the theory theorist's account.

Two such virtues seem worth considering. The first is simplicity. Earlier I conceded simplicity can sometimes outweigh depth as an explanatory virtue. A defender of the theory view might insist that even if both a theory theorist account and a simulationist account of mental state ascription were true, the theory-theorists account was simpler.

It is not obvious that this is true. However, instead of entering into a discussion of the relative simplicity of the two accounts, the defender of simulationism should just deny that simplicity is anything other than marginally relevant in this context. It seems plausible that simplicity is primarily a pragmatic virtue: simpler explanations are easier to understand, assimilate, and communicate and so on. If so, then to the extent that our primary interest in understanding how the mind works, we should allow simplicity to be outweighed by explanatory depth when assessing alternative explanations.¹⁷

A second explanatory virtue that the theory theorist might appeal to is theoretical unification.¹⁸ On this line, it would be accounted a virtue of the theory theorist's account that it appealed to the same sorts of theoretical notion as the ones which are

¹⁶ Putnam (1975).

¹⁷ It's been suggested to me (by Rachel Cooper) in discussion that in a case like this we may simply want to accept that the two explanations have different, competing explanatory virtues and that both explanatory programs are worth pursuing for different purposes. I think that this would be enough for most proponents of simulationism. But it's far from clear to me that the theory view really does score over the simulation view in respect of any significant explanatory virtue.

¹⁸ Friedman (1974), Kitcher (1981).

appealed to in cognitive scientists' explanations of other human capacities, while the simulationist' account does not.

Explanatory unification can be valuable for different reasons. Sometimes it makes for simplicity in an explanation. I have argued that we should discount simplicity. Explanatory unification can also be valuable in helping to 'carve nature at the joints'. However, explanatory unification can be a dangerous virtue. For unified explanations can sometimes encourage us to carve in the wrong place.

This is exactly the point that the simulation theorist should make in the face of a combined appeal by the theory theorist to Collapse and the greater unification promised by the theory theory. If the simulationist explanation of our folk psychological capacities is correct, then even if the Collapse argument goes through, the theory-theorists account of our capacities is misleading because it encourages us to assimilate our folk psychological capacities to other capacities that we can account for in terms of our tacit knowledge of a psychological theory. For what the truth of the simulationist explanation would show is that our folk psychological capacities are importantly different from these other capacities because of the way in which they depend primarily on a redeployment of other cognitive resources.

6 A Final Loose End

I have argued that a proponent of simulationism should respond to Collapse by appealing to considerations of explanatory depth. However, one might feel that this can yield only a Pyrrhic victory. For, on the account of explanatory depth that I have relied on, an account of our capacity for folk psychological explanation in purely neurological terms would have greater explanatory depth than either the theory theorist's or the simulation theorist's account, and would therefore have to be preferred to either of them. My response here is to bite the bullet. Maybe an explanation of our folk psychological capacities in neurological terms would have greater explanatory depth than a simulationist account. But we are still a very long way from having such an account. Until we do, the simulationist account holds its own—at least as far as the Collapse is concerned.

A further possibility which deserves discussion has been mooted by Alvin Goldman, who has suggested that a simulationist may be able to take a relaxed view about how a capacity to simulate is implemented, and may want to allow for the possibility of a capacity to simulate being by possession of a tacit theory.¹⁹ For the purposes of this paper I do not want to debate the merits of this suggestion—call it *Implementation*, but simply to consider whether it undermines my argument against Collapse. I do not think that it does.

The objection to my argument would be that if Implementation is correct, this undermines the suggestion that Collapse entails that there is an explanatory asymmetry between the simulationist account of mental state ascription and the theory theorist's account of the sort which supports a claim of greater explanatory

¹⁹ Goldman (2006). Mazviita Chirimuuta raised this issue with me in discussion at an Open Session of the Joint Session of the Mind Association and Aristotelian Society at Bristol in 2007.

depth for the latter. The idea would be that if a capacity to simulate is implemented by a theory, then the correctness of an explanation of our capacity to mentalise at the implementational level would also explain the correctness of the simulationist account. If so we could no longer claim that the simulationist account was explanatorily deeper.

Goldman has not worked out the suggestion that a capacity to simulate might be implemented by our possession of a theory in any detail, and it is not clear that this idea is consistent with the details of the neurologically based story that he wants to tell about some forms of simulation-based mentalising (such as the attribution of emotions). But leaving this on one side, one response that the opponent of Collapse is entitled to is as follows. A story about how a capacity to simulate was implemented by a tacit theory might tell us how simulation was possible. But it would not, by itself, show that a theory-based explanation satisfied the criterion for explanatory depth that I articulated above. Here's why.

The criterion I articulated said that one explanation was deeper than another if the first explanation told us why the second explanation is correct, and not vice versa. For Implementation to defeat the claim that the simulationist explanation is deeper than the theory theorists, it has to be the case that it explains why the theory view is correct. However, it is a significant part of the simulationist's explanation of mentalising that mentalising involves the redeployment of capacities that we use in a first order way in other contexts. So in order to satisfy the criterion, the story about how simulation was implemented by grasp of a tacit theory would have to explain this.

Given the somewhat sketchy nature of the proposal that Goldman suggests, it would be hard to argue that no story about implementation could do this. However, it does seem fair to conclude that there is little reason to think that a story about the implementation of a capacity to simulate by possession of a tacit theory would *have* to do this. One reason for being skeptical about this possibility would be this: why would an implementational explanation of our capacity to simulate even entail that we had the sorts of further capacities involved in first order deployment of concepts, let alone tell us anything about how those capacities related to the capacities implemented by the tacit theory?

If so, then at very least the implementation proposal requires further development before it gives us any reason for being skeptical about my argument against Collapse; and there are some reasons for thinking that an explanation of this sort could not be given. This of course does not show that Implementation could not be true; but only that its truth would provides little support for Collapse.

Acknowledgements I am grateful to Jane Heal, Adam Morton and Radu Bogdan for advice and encouragement, and to Josh Cowley, Hilmi Demir, Max de Gaynesford and Mark Ashton Smith and two referees for *Erkenntnis* for useful discussion and feedback.

References

- Carruthers, P. (1996). Simulation and self-knowledge: A defense of the theory theory. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind*. Cambridge: Cambridge University Press.
- Davies, M. (1987). Tacit knowledge and semantic theory: Can a five percent difference matter. *Mind*, 96, 441–462. doi:[10.1093/mind/XCVI.384.441](https://doi.org/10.1093/mind/XCVI.384.441).

- Davies, M. (1994). The mental simulation debate. In C. Peacocke (Ed.), *Objectivity, simulation and the unity of consciousness*. New York: Oxford University Press.
- Davies, M., & Stone, T. (2001). Mental simulation, tacit theory and the threat of collapse. *Philosophical Topics*, 29, 27–173.
- Friedman, M. (1974). Explanation and scientific understanding. *The Journal of Philosophy*, 71, 5–19. doi: [10.2307/2024924](https://doi.org/10.2307/2024924).
- Goldman, A. (2006). *Simulating minds: The philosophy, psychology and neuroscience of mind-reading*. Oxford: Oxford University Press.
- Heal, J. (1994). Simulation vs theory theory: What is at issue? In C. Peacocke (Ed.), *Objectivity, simulation and the unity of consciousness*. New York: Oxford University Press.
- Heal, J. (1996a). Simulation, theory and content. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories of mind* Cambridge. Cambridge: Cambridge University Press.
- Heal, J. (1996b). Review: Belief, simulation and the first person: Comments on a study of concepts by Christopher Peacocke. *Philosophy and Phenomenological Research*, 56, 413–417. doi: [10.2307/2108532](https://doi.org/10.2307/2108532).
- Heal, J. (1998a). Understanding other minds from the inside. In A. O'Hear (Ed.), *Current issues in the philosophy of mind* (pp. 83–99). Cambridge: Cambridge University Press.
- Heal, J. (1998b). Co-cognition and off-line simulation. *Mind and Language*, 13, 477–498. doi: [10.1111/1468-0017.00088](https://doi.org/10.1111/1468-0017.00088).
- Hutto, D. (2004). The limits of spectatorial folk psychology. *Mind and Language*, 19, 548–573. doi: [10.1111/j.0268-1064.2004.00272.x](https://doi.org/10.1111/j.0268-1064.2004.00272.x).
- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48, 507–531. doi: [10.1086/289019](https://doi.org/10.1086/289019).
- Morton, A. (1996). Folk psychology is not a predictive device. *Mind*, 105, 119–137. doi: [10.1093/mind/105.417.119](https://doi.org/10.1093/mind/105.417.119).
- Morton, A. (2003). *The importance of being understood: Folk psychology as ethics*. London: Routledge.
- Nicholls, S., & Stich, S. (2003). *Mindreading—An integrated account of pretence, self awareness and understanding other minds*. Oxford: Clarendon Press.
- Peacocke, C. (1992). *A study of concepts*. Cambridge, MA: MIT Press.
- Putnam, H. (1975). Philosophy and our mental life. *Mind, Language, and Reality*. Cambridge: Cambridge University Press.
- Ravenscroft, I. (2003). Simulation, collapse and human motivation. *Mind and Language*, 18(2), 162–174. doi: [10.1111/1468-0017.00219](https://doi.org/10.1111/1468-0017.00219).
- Wringe, B. (2003). Simulation, co-cognition and the attribution of emotional states. *European Journal of Philosophy*, 11, 354–373. doi: [10.1111/1468-0378.00190](https://doi.org/10.1111/1468-0378.00190).