BJDP 625—14/4/2009—NAGARAJAN—334906

British Journal of Developmental Psychology (2009), 00, 1–6
© 2009 The British Psychological Society

The British Psychological Society

www.bpsjournals.co.uk

Commentary

Is there an alternative to simulation and theory in understanding the mind?

Peter Mitchell[1]*, Gregory Currie[2] and Fenja Ziegler[1]

[1]School of Psychology, University of Nottingham, Nottingham, UK
[2]Department of Philosophy, University of Nottingham, Nottingham, UK

We thank the commentators for their reflections on our article, reflections that have advanced our own thinking on this intriguing topic. Of our two commentators, Apperly is the more critical, challenging not only our approach but the framework in which it is embedded; Harris's comments are broadly supportive, while making it clear that there is no room for complacency in defending our account. We begin with comments on two suggestions of Harris and we then turn to an issue raised by both Harris and by Apperly concerning mind reading in infants. The final four sections are devoted to responding to criticisms from Apperly.

Two open questions

Harris makes an important suggestion about something our approach to the development of mind reading might explain: the puzzling fact that children who pass false belief tests will often fail to comprehend that the agent in question will, as a consequence, experience a certain emotion. Instead, the children fall back on the default attribution: they think that Little Red Riding Hood (RRH) will be afraid, even though they think she does not know that it is the wolf dressed as her grandmother. As Harris points out this cannot be simply because the children find the downstream effects of belief harder to understand than the belief itself, because children who pass false belief tests also understand the likely behaviours that flow from the belief – looking in the wrong box for the sweet. Harris suggests that the true explanation is that emotions are more salient than belief and that the dominance of the child's own emotion – fear of the wolf – defeats the child's capacity to override her own mental state.

This seems to us very plausible, and Harris nicely illustrates how the theory might be tested by invoking an emotional stimulus with lower salience. One thing calls for clarification: when the child contemplates RRHs predicament – she thinks, wrongly, that it is her grandmother in the cottage when in fact it is a very dangerous creature – what emotion does this evoke in the child? Is it fear of the wolf, or fear for RRH.

* *Correspondence should be addressed to Professor Peter Mitchell, School of Psychology, University of Nottingham, University Park, Nottingham NG7 2RD, UK (e-mail: peter.mitchell@nottingham.ac.uk).*

DOI:10.1348/026151009X441935

Philosophers have often complained that empathy-based accounts of our relations to fictional things do not distinguish sufficiently between these possibilities, the worry would be that, since, the child presumably feels afraid for RRH, and this emotion is highly salient to the child, our account would end up predicting that the child will attribute to RRH the emotion *feeling afraid for herself*, which certainly does not seem to happen. To this a number of replies are possible. It might be argued first that young children, while aware at some level that the story is a fiction, do actually feel afraid of the wolf, though they may in addition (and indeed as a consequence) feel afraid for RRH. Secondly, young children may not be able to identify their own emotional states to this level of refinement, the child who actually feels afraid for RRH may interpret this as fear of the wolf, and make an attribution accordingly. Or it may simply be that the child experiences, and attributes to the character, a generalized emotion – fear – without further discriminating characteristics. Perhaps there is an interesting developmental story here: do children go through a period when they respond to the perceived dangers of others simply by feeling the emotion that the person would feel if they knew of the danger, and then shift at some point to undergoing the distinct, 'third-person' emotion of fearing for the character concerned?

Harris makes another important and this time cautionary point: even very young children are able with ease to set aside current reality in the context of a story, they cope easily with stories of talking animals, and even catch the spirit of the enterprize, attributing speech and understanding to animal characters where there is no explicit direction in the story to do so, but where that is the natural generalization from what the story says. So we must be careful not to attribute to young children a general incompetence in setting aside current reality, as they understand it. How should this affect our approach to the explanation of children's developing understanding of mind, and especially of the contrasting view points of others?

We do not see – and Harris does not suggest – any outright contradiction between the ideas that (1) the development of mind reading is a story of progressive mental flexibility and improving capacity to set aside current reality and (2) very young children are able to catch hold of a rule or principle implicit in a fiction according to which massive violations of current reality are in force. But it does look as if any explanation of their compatibility will have to place heavy weight on the idea that children find it easy to set aside current reality when prompted to do so by a story and find it very hard to do so when those prompts are not in place. The rather thinly described narrative presented in a standard false belief task does not ask the child to set aside any assumptions the child would naturally make in comprehending the situation presented, and therefore, the child's default assumptions govern her understanding–as they would, presumably, in a story where there are people and animals but no indication in the story that the animals have human-like characteristics (though there is an extra difficulty here that young children may be so used to stories with talking animals that they simply assume that any story with animals in it belongs to that genre). What explains the power of narrative to shift young children from their default assumptions? An account of this might appeal to the important evolutionary role of narrative discourse in knowledge transmission, securing common social purposes and maintaining honest signalling (see Currie, 2009). But such an account would have to explain the shear profligacy of our capacity to adjust defaults in response to narrative. It's useful to us to understand stories told from perspectives which differ in various ways from our own current perspective, because that helps us understand both the teller and the world better. But why are we so easily captivated by impossible tales in which animals talk or people travel through time to become their own grandparents?

BJDP 625—14/4/2009—NAGARAJAN—334906

Are infants different from older participants in their ability to mentalize?

In our paper, we cited evidence from Onishi and Baillargeon (2005) who found that infants expected an actor to act according to his or her belief, even when that belief contradicted the infant's own, we cited this in support of the view that even very young children have a basic competence in understanding beliefs. Comments from both Apperly and Harris have allowed us to see that the issue here is problematic.

What does infant competence, as revealed by looking paradigms, tell us about simulation? One can take any of several views here. The first would be to say that infants are competent simulators but that simulative competence declines as language develops, picking up again only around age four. Against this, there is evidence, which we cited, that 3-year-olds who fail false belief tests behave in such a way as to indicate that, at some level, they know the right answer. This suggests an alternative hypothesis: that children from a very young age onwards have some grasp of the way in which a false belief will guide an agent's behaviour, but that simulative methods are required in order to bring this grasp to the level of articulation. There is in fact some independent evidence to support this hypothesis. It has been believed for some time that people make judgements of handedness – is it a left or a right hand being displayed? – by engaging in simulated hand movement, mentally rotating their own hands from their current actual position into a position congruent with that displayed in the picture (e.g. Parsons, 1987). But when people report the episodes of motor imagery involved, they rarely report first of all moving the incorrect hand, it seems therefore that, at some level, a decision about whether the hand is left or right has occurred before the mental movement. However, it should not be concluded from this that the mental movement is irrelevant to enabling the participant to give the right answer. Experiments with patients who have undergone hemispheric separation indicate that when the patient is presented with a left hand in the left visual field the patient is able reliably to identify it as the left hand, the left hand being, in effect, presented to the right hemisphere, which controls action of the left hand and is hence able to simulate left hand movement. But when a right hand is presented to the left visual field and hence to the right hemisphere, performance is at chance, the hemispheric disconnection means that the left hemisphere cannot now be recruited to the simulation task (Parsons, Gabrieli, Phelps, & Gazzaniga, 1998). The reasonable conclusion from this is that simulated hand movements function to give the subject access to implicit knowledge of the correct answer to the question 'Which hand is being presented?' It is possible, by parallel reasoning, to conclude that simulation in false belief tasks gives the subject access to a pre-conceptual grasp of the idea that the target agent will act on a false belief. This is, we stress, speculation, we grant that this is a difficult issue, and thank our commentators for raising it.

Is there a link between counterfactual reasoning and mentalizing?

Apperly suggests that, we too easily link counterfactual and false belief reasoning as the products of the same, simulationist mechanism. In fact, he suggests, there are crucially different notions of simulation involved here: theory- and process-driven simulations. The former drives counterfactual reasoning while the latter drives false belief reasoning. We disagree.

When undertaking a false belief task one reasons, not from what you believe, but from what the target agent believes. Thus, an observer in a standard false belief task, asked to predict Maxi's behaviour, might reason from the premize: 'The last time I saw it, the sweet was in box A' to the conclusion 'The sweet is still in box A'. The observer does not believe

the premize of the argument; from her point of view, it is a contrary-to-fact assumption. But it is the right assumption to reason from because it is what the target believes. In reasoning this way, the observer treats the premize as if she did believe it, reasoning from it in just the way she would if she did believe it. She simulates the state of one who does believe it. And in doing so, she draws on things she does believe, such as the proposition that, by and large, things tend to stay in the same place, at least over short periods of time, that inanimate things such as sweets do not move of their own accord.

Suppose now I am faced with a non- (theory-of-mind) ToM task: I want to decide whether a counterfactual is true. The counterfactual in question is, let us say, 'If Caesar had been in charge of UN forces in Korea he would have used the atom bomb'. This statement invites me to imagine something: Caesar being in charge of UN troops in Korea. I am then to see what conclusion, if any, this imagining gets me to concerning how the war would have been conducted.

Two points are worth making about this. First, in doing this, I am simulating the state of one who believes Caesar is in charge in Korea. There may not actually be anyone who believes this, and so to this extent this case is different from the previous one. But this cannot be grounds for saying that the second case is not simulation, or at least not simulation of the same kind as in the first case. There is such a thing as failed simulation: I try to create imaginative versions of your beliefs with a view to working out what you will do, but fail to identify the beliefs you actually have. The beliefs I simulate are not the beliefs of my target, but I am simulating nonetheless.

The second point is that, in the case of Caesar as in the case of the sweet, the inference depends on knowledge or at least on belief. If I know or believe nothing about Caesar (about his personality, attitudes to fighting a war, etc.), I will not be able to draw any relevant conclusion. In particular, I shall be at a loss to know whether he would have used the atom bomb. Again there is a difference: the knowledge needed to make a judgment about Caesar is rather specialized, whereas the knowledge needed to infer that the sweet remains in box A is extremely general. This cannot be a reason for saying, once again, that the second case is not a case of simulation, or not simulation of the same kind as in the first case. Suppose, the false belief task had been somewhat more difficult: there is a witch who moves sweets from one place to another unless she happens to be absent that day, there is also a variety of puppets hanging around in various costumes. To predict where Maxi will look for the sweet now, I have to do a simulation of reasoning that depends on bringing to the task knowledge about what witches look like, so as to be able to tell whether one is in the room or not. That knowledge is fairly specialized, though perhaps not so specialized as knowledge of Caesar's career as a general, but the difference here is only one of degree.

In both the false belief task and the Caesar case I am solving problems by process-driven simulation. In both cases the simulations draw on knowledge, that is perfectly legitimate, because the simulations in question are simulations of reasoning, whereby a person uses his or her reasoning processes to simulate another, actual or hypothetical, piece of reasoning. Thus, the processes that drive the simulations are processes of reasoning, and reasoning is a process that draws on knowledge. Simulations which depend on knowledge can be defined as *theory-driven*, but this will not deliver the conclusion Apperly wants: that counterfactual reasoning is typically done by some process of a different kind from the process that helps us to perform ToM tasks, for both tasks will now turn out to involve theory-driven simulation.

Of course there are differences between the counterfactual task and the ToM task. In the counterfactual task one ends up assessing the truth-value of a counterfactual, and

in the ToM task one ends up predicting what someone will do. When it is claimed that they are both tasks which depend on the same facility with simulation, what is being claimed is that both tasks involve simulation – and the same kind of simulation. It is not being claimed that this is all that they involve.

Does the gradual acquisition of multiple rules explain children's gradual improvement in acknowledging false belief?

Apperly challenges our suggestion that gradual improvement in children's success in acknowledging false belief selectively favours a bias-simulation account. He suggests that if children gradually acquired multiple rules for mentalizing, that too would lead to gradual improvement in correct attributions of false belief. Is that true? It depends, we claim, on the particulars of the rules in question. It may be possible to postulate a set of rules, along with a timetable for their acquisition, which is consistent with gradual improvement, so far as we know, no one has done this, so whether such a set of rules and corresponding timetable would have independent plausibility is difficult to say. But certainly, not just any set of rules will do, some rules will be such that they assist performance of false belief tasks only in conjunction with other rules, thereby predicting a step-change in performance as the whole set is acquired. Apperly does not indicate what set of rules he has in mind, though he uses as examples rules which serve to supplement the rule (R) 'If agent A was present when fact X was manifest then A knows about X'. These supplementary rules are that A should be 'sentient', should be 'attentive', and should 'know enough already to grasp X when X was made manifest'. These seem to us more like statements of primitive capacities rather than acquired rules, but assume for the moment that they are rules. If all of these rules are needed to perform well on a false belief task, then they would all be needed together, and their serial acquisition would not account for improving performance. Suppose a child has acquired only the first rule (R), but not any of the others. How would that explain getting the answer on a false belief task right part of the time? The rule about sentience does not add anything, since sentience is part of what it takes to be an agent; the rule about attentiveness also does not seem to add anything relevant, since false belief tests do not differ in the amount of attentiveness the puppet characters display; similarly for the final rule, since these tests do not vary in what the puppet characters understand about what is manifest. We repeat: it may be possible to find a set of rules which fits the bill, but any such proposed set will have to be independently justified. By contrast, the simulation theorist has an entirely principled reason for expecting performance to improve. The basis of the simulationist idea is that, we exploit our own mental similarity to others in order to understand them, we could not do this unless there was a presumption in favour of the idea that they are mentally like us – a presumption from which we shift only when there is reason to do so. It is to be expected, therefore, that mental adjustment to fit another person's reasoning will be effortful, and hence subject to improvement.

Is it unreasonable to expect Wellman et al.'s meta-analysis to identify a U-shaped developmental function?

Apperly suggests that due to exclusions of children who made errors on control questions, the sample examined by Wellman, Cross, and Watson (2001) does not include the very children who genuinely responded in the most primitive way, which is with equal probability to the true belief and false belief locations in an unexpected transfer test of false belief. Wellman *et al.* (2001) included 77 articles, reporting

178 separate studies with 591 conditions. Some conditions were excluded because of comparability issues, atypical samples or other issues unrelated to Apperly's point, but only nine conditions were excluded because of errors relating to control questions (fewer than 60% of children answered the control questions incorrectly or 40% of the participants were dropped). We therefore, remain confident that Wellman *et al.*'s analysis includes the children relevant to our analysis.

Besides, the point we intended to make is that in principle Wellman *et al.* do not recognize two patterns of incorrect responding: One pattern is to select randomly between two locations in a test of false belief. The other pattern is to select the true belief location systematically, errors that are not explained or even recognized by Wellman *et al.* In identifying the two patterns of incorrect responding, we pose a further question concerning their developmental sequence and it is this latter question that ought to be of special import to Wellman *et al.*'s account.

Should we use the term 'imagination' instead of 'simulation'?

Apperly questions not only the specific proposals, we make but the framework within which our discussion takes place. While our paper is an attempt to get beyond a stark opposition between simulation and theory, we do assume that the best proposal is likely to be a combination of these approaches. Apperly, on the other hand, wonders 'whether they are games worth playing at all'. Some of our earlier comments in this reply will, we hope, have added to the case for thinking that they are. But there is something in Apperly's sub-title that deserves to be addressed: 'Imagination and rule-use may be better than simulation and theorising'. We see no contrast between these pairs. Simulation theory has always been an attempt to make more precize the traditional but vague idea that understanding others is an imaginative task, involving projection into the situation of another. And understanding mind reading as involving rule acquisition and use is one version of the view that mind reading involves theory. We do not reject the possibility that there is a better approach, but we also remain unclear on what Apperly thinks this approach is.

In the light of our response to the commentaries, we feel that, at present, there is no alternative to simulation and theory in understanding the mind.

References

Currie, G. (2009). *Narratives and narrators: A philosophy of stories*. Oxford: Oxford University Press.

Onishi, K. H., & Bailargeon, R. (2005). Do 15-month-old infants understand false beliefs? *Science*, *308*(5719), 255–258.

Parsons, L. M. (1987). Imagined spatial transformations of ones hands and feet. *Cognitive Psychology*, *19*(2), 178–241.

Parsons, L. M., Gabrieli, J. D. E., Phelps, E. A., & Gazzaniga, M. S. (1998). Cerebrally lateralized mental representations of hand shape and movement. *Journal of Neuroscience*, *18*(16), 6539–6548.

Wellman, H. M., Cross, D., & Watson, J. (2001). Meta-analysis of theory-of-mind development: The truth about false belief. *Child Development*, *72*(3), 655–684.

Received 6 March 2009; revised version received 1 April 2009