BJDP 415—3/7/2008—ROBINSON—302742

British Journal of Developmental Psychology (2008), in press

The British Psychological Society

www.bpsjournals.co.uk

Two routes to perspective: Simulation and rule-use as approaches to mentalizing

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 review evidence relating to children's ability to acknowledge false beliefs within a simulation account according to which our focus is set by default to the world as we know it: hence, our current beliefs assume salience over beliefs that do not fall into this category. The model proposes that the ease with which we imaginatively shift from this default depends on the salience of our current belief, relative to the salience of the belief that is being simulated. However, children do use a rule-based approach for mentalizing in some contexts, which has the advantage of protecting them from the salience of their own belief. Rule-based mentalizing judgements might be faster, cognitively easier and less prone to error, relative to simulation-based judgements that are much influenced by salience. We propose that although simulation is primary, rule-based approaches develop as a shortcut; we thus grow from individuals capable of using only simulation into individuals capable of both techniques.

It can seem mysterious as to how anyone can know what is in the mind of another; indeed, philosophers used to discuss the question in the style of someone responding to a sceptical challenge: how is it even possible to know another mind? The cognitive revolution in psychology brought forth a more positive approach: mental states are inferred from behaviour (including speech) just as the unobservable particles and forces of science are inferred from the observables. This makes knowledge of other minds knowledge by theorizing or, as it is sometimes said, by appeal to rules. Developments in the 1980s – some of them from philosophers interested in the sciences of mind – suggested an apparently very different answer. On this view we use our capacity to imagine ourselves in the other's situation and then simply note what thoughts or decisions we have, something that can be done without having a theory about the relations between mind and behaviour.

If either of these approaches – now commonly called theory–theory and simulation theory – is to be credible, it must do more than sketch an in-principle answer to the question 'How do I know what you are thinking?' It must tell a testable story about how

* *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:10.1348/026151008X334737

children get to the point where they can access the point of view of another person. No satisfactory story of this kind is currently available.

We might try to supply one by taking sides, asking: does the development of mind-reading skills in children depend on a grasp of rules, or – as simulation theorists have claimed – on an increasingly flexible capacity to project one's self imaginatively into the position of another agent? These approaches to the question have sometimes been regarded as at odds with one another and in extreme formulations they are incompatible. We suggest, however, that a plausible account will appeal to a mix of strategies. The suggestion that mind-reading involves some mix of theory and simulation has appealed to authors in the past (Gopnik & Wellman, 1992; Perner, 1995), but this did not lead to the formulation of a general theory simulation account of the development of false belief understanding or an account of the relative contributions to particular tasks of simulation and theory. The challenge of formulating a hybrid account has been picked up in more recent publications, which continue the trend of viewing rule-based and simulation accounts not as mutually exclusive, but as working together in some form of hybrid (Goldman, 2006; Nichols & Stich, 2003). These existing accounts are mainly philosophical and theoretical in nature and do not attempt to embed a hybrid account within the vast empirical literature on theory of mind which spans the last 25 years. The plethora of empirical data, and seemingly contradictory evidence, keeps on being enriched by new insights, which complements some assumptions on mentalizing but contradicts others.

The challenge then remains to develop a theory that describes the circumstances under which simulation or a rule-based approach is more likely to be used. Accordingly, we identify task demands and developmental factors that may be critical in determining whether simulation or a rule-based approach is employed in a particular context. In doing so, we will integrate empirical research from the last 25 years, moving from classic research to recent findings to formulate a model based on an extensive review of available empirical evidence. Amongst the recent findings we will explore the possibility that very young children show an understanding that people's actions are guided by their beliefs (e.g. Onishi & Bailargeon, 2005), we will look at a possible account for a neural basis of mentalizing described in the mirror neuron network and we will examine accounts of adults' continuing problems with specific mentalizing tasks. In doing so, we will engage with classic and disputed evidence concerning the development of theory of mind abilities, in recognizing that a successful hybrid model needs to be able to give an account based on a broad and deep range of empirical findings. We begin with a brief account of the contrast between simulation and rule-based judgments.

Contrasting the simulation and rule-based stance

Intuitively, the difference between simulation and rule-following[1] is a difference between stances we take towards other agents. When we adopt a rule-based approach we treat agents as objects of investigation, much as when we investigate the behaviour of planets and electrons: we look for rules and initial conditions from which we can

[1] *There are of course many versions of theory–theory and simulation theory and what we define here are theoretical positions which fall into generally acceptable parameters of these positions. We do not feel bound to embrace either radical simulation (e.g. Gordon, 1996), or radical theory–theory (e.g. Gopnik & Wellman, 1992).*

predict or explain their behaviour, though the rules we use in the case of agency may be different in kind from those we look for in the case of planets and electrons. When we adopt the simulative stance, in contrast, we seek to place ourselves imaginatively in the position of the agent, and proceed as if our own mental processes will operate in ways that are roughly congruent with those of the target agent. Consider an analogy: I want to know how fast your car will go up this hill. One way to find out would be to formulate a theory about the car and the road in question, the power of the engine, the way this power is delivered to the wheels, the resistance offered by the surface, and so on. Few of us have ready access to such a theory, and if we did, the task of deriving the right consequences would be long, tedious, and prone to error. If I happen to own a car of the same model and year as yours, quite a good way to answer the question would be to drive up the hill as fast as I can. The information gained from the simulation may not be wholly reliable but will probably be more reliable than any conclusions I manage to wring from the relevant theories of cars and roads.

Simulations can themselves be rule-based. If you are wondering what is wrong with your car you may draw on such rules of thumb as 'If the engine won't start that may be because the carburettor is flooded', and consequently start investigating the state of the carburettor. Suppose I want to predict your behaviour in response to the fault, and so try to simulate your reasoning; I now have to go through the same rule-based reasoning. How, then, can there be a contrast between simulation and the use of rules? In part, the mind-reading debate has been over the use of psychological rules, and simulation theorists have been sceptical about the usefulness or availability of such rules. To understand your behaviour in response to the car breakdown I do not have to have access to a rule which says 'When people's cars break down they think about whether the carburettor is flooded.' Instead I put myself in your situation, and think about carburettors. I am much more likely to have a rule concerning carburettors than I am to have a rule about thoughts about carburettors, and if I do not have the first I cannot have the second (Heal, 1996). Thus simulation, even when it involves rule-use, may enable us to predict behaviour without appealing to possibly complex psychological generalizations.

Even so, simulation might not always be the best method for mentalizing. If a particular problem is familiar, allowing the initial conditions to be identified with ease, and the rule for predicting behaviour of an agent is readily available, then a rule-based solution might be quick, relatively effortless and tolerably accurate. Suppose we observe that a person did not witness his chocolate being moved from location A to B: this is a familiar initial condition and we can invoke the rule that the person will retain his/her initial outdated belief that his/her chocolate is in A where he/she last saw it. Accordingly we expect him/her to embark on a vain search in A. In this case, it may not be necessary to simulate the person's mental states. In other cases, however, the initial conditions may be unfamiliar, where seeking a relevant rule could be arduous and time-consuming; simulation might be the more effective strategy.

However, the theory we shall propose is not the comfortably ecumenical one that children depend on simulation and rule-following in about equal measure, with no priority for one over the other. While allowing for occasions of rule-based reasoning, we argue that the underlying picture of children's developing skills in mind reading is one of gradual change. Rule-based accounts have difficulty with gradual change because the transition to competence should be sudden if it depends on acquiring a rule. The better explanation is that children improve with age in their ability to set aside their own current and more salient beliefs in favour of a hypothetical alternative (Harris, 1991). This, in turn, is best accommodated by the simulation approach. Such an account

explains why children start out systematically reporting their own current beliefs; it also explains the developmental move away from this without needing to predict (or explain) a sharp transition to a state of competence.

We begin by reviewing rule-based approaches, which we claim fail to accommodate the big developmental picture.

Setting out the debate: Children's performance on tests of false belief

In the classic unexpected transfer test of false belief understanding (Wimmer & Perner, 1983), Maxi puts his chocolate in the blue cupboard, then leaves. Unknown to him, it is subsequently moved to the red cupboard. Observing child participants then predict where Maxi will look for his chocolate. Rule-theorists claim that when participants correctly predict that Maxi will look in the blue cupboard, they effectively acknowledge that Maxi's belief was formed on the basis of his seeing the chocolate in the blue cupboard; Maxi did not see the chocolate in the red cupboard and therefore his search will be based on his outdated information. A child who wrongly predicts that Maxi will search in the red cupboard effectively fails to appreciate that Maxi's lack of informational access leads him to act upon an outdated belief. In general, supposedly, young children struggle to acknowledge false belief because they do not understand the rule which connects informational access and the consequent state of knowledge, or, more specifically, between seeing and believing and its converse: if you do not see then you do not know (Wimmer & Gschaider, 2001; Wimmer, Hogrefe, & Perner, 1988; Wimmer & Weichbold, 1994).

Independent evidence suggesting that young children fail to understand the relation between information and knowledge emerged in various studies. Wimmer *et al.* (1988) found that while children aged about 3 years were accurate in reporting whether or not a person had seen an event, they were less effective in reporting whether or not the person knew about the event in question. Older children made equally accurate judgements about seeing and knowing, and there was a strong correlation between the two kinds of judgment. Gopnik and Graf (1988) found that children aged about 3 years seemed oblivious to how they came to know a fact, again suggesting they did not make a connection between information and knowledge.

Children whose natural language is English seem to learn that the past tense of a verb is formed by adding – ed, and then initially apply the rule too generally, transforming irregular verbs into regular verbs: as with 'I runned'.[2] If mentalizing is rule-based, then we might expect to find similar evidence of over-generalization, and a study by Sodian and Wimmer (1987) yielded circumstantial evidence in support of this possibility. Children were shown a choconuts bag and then witnessed the experimenter take one of the things (they could not see what, but presumably it was a choconut) from the bag and transfer it to a box. Children aged about 5 years had no difficulty inferring that the thing in the box was a choconut, and yet denied that another person, with precisely the same information, would know there was a choconut in the box. Paradoxically, even though children could make an inference in this case, their denial that another person knew the content of the box seemed to suggest that they did not understand that the process of inference could serve as a way of gaining information; seemingly, they had gained knowledge without knowing how.

[2] *Note that Rumelhart and McClelland (1986; also, Plunkett & Marchman, 1991) used a connectionist network to explain this phenomenon not as over-application of a rule but as over-attention to a statistical regularity.*

According to Sodian and Wimmer (1987), children's denial of the other person's knowledge was based on the fact that the other had not looked inside the box. Children seemed to be applying the 'don't see–don't know' rule too generally, and were led thereby to deny that a person could know a fact via another route, such as inference. Sodian and Wimmer call this 'inference neglect' (see Rai & Mitchell, 2006, for an investigation into the scope of this phenomenon).

Do young children fail tests of false belief because they lack a rule for connecting information with knowledge? Apparently not: there is ample evidence that children understand the relationship between seeing and knowing well before they are typically able to pass false belief tests. As a case in point, Robinson and Mitchell (1995) presented a scenario about identical twins. Initially, both placed a ball in the blue drawer then left the scene. Shortly after, one of the twins returned (we do not know which because of their identical appearance), transferred the ball to the red drawer, and left. Finally, both twins returned to the scene and Mother asked them to fetch the ball. One twin went promptly to the blue drawer, where they had jointly put the ball first of all, and the other went to the red drawer, where one of them hid the ball subsequently. Observing child participants were now asked which twin had stayed outside: the one who went to the ball's current location (red), or the one who went to its first location (blue). Seemingly, quite a sophisticated inference is required to work out that the twin who stayed outside was the one who went to the currently empty blue drawer, yet a remarkably large number of children aged around 3 years succeeded. In Robinson and Mitchell's investigation 1, 85% of 3-year old gave a correct judgment, while only 30% were correct in predicting which drawer the absent twin would go to – a condition similar to a standard unexpected transfer test of false belief. The findings suggest a surprisingly early ability to link informational access with the consequential state of knowledge.

Early understanding also emerged in a study by Robinson, Champion, and Mitchell (1999; also, see Robinson & Whitcombe, 2003). Children participated in a game in which they had to state the content of a box. Their initial statement was contradicted by another person and then children had to make a final statement in which they either reaffirmed their initial statement or shifted to agree with the other person. In one condition, children looked into the box and saw the content before making their initial statement, while in another condition, only the other person looked into the box. Even the youngest children, aged about 3 years, tended to change their statement to agree with the other person when he (but not they) had looked into the box, but reaffirmed their initial statement when only they themselves had looked. In other words, children made a sophisticated connection between the veracity of a statement and informational access: someone who has visual access to a fact is equipped to make a well-informed statement and should be believed. These findings suggest that the ability to link information with knowledge is acquired before children begin making correct judgements on a standard test of false belief.

Perhaps a rule-based account would remain useful if it were modified to posit that possessing a rule linking seeing and knowing is necessary but not sufficient for passing tests of false belief. In that case we need an account of what other factors are involved. If these other factors are not themselves rules, then we do not have a fully rule-based account.

Another difficulty for accounts that appeal to rules of any kind is that children's progress towards reliably passing false belief tests seems to be gradual. A meta-analysis conducted by Wellman, Cross, and Watson (2001) suggests that children start out systematically giving incorrect judgements. When they are a little older, they give a

mixture of correct and incorrect judgements, and finally, when older still, they give systematically correct judgements. This description of development, especially at its intermediate point, is supported by studies that use a test–retest paradigm. Mayes, Klin, Tercyak, Cicchetti, and Cohen (1996) found a lack of consistency over repeated presentations of a test of false belief. Surprisingly, some children who gave a correct judgment at initial testing went on to make an incorrect judgment 6 weeks later. Hughes *et al.* (2000) found better consistency over different testings, but the absolute level of consistency was still less than impressive. Furthermore, longitudinal studies reveal that the number of tests of false belief that children pass increases very gradually with age (Amsterlaw & Wellman, 2006; Flynn, O'Malley, & Wood, 2004). These findings suggest that developmental change is not as sharp as would be expected if children had acquired a rule for inferring beliefs.

It might be argued that the data are consistent with a rule-acquisition account if we assume that children, like scientists, undergo a process of 'progressive theory replacement'. In science it is rare for a theoretical breakthrough to result in a giant leap in explanatory power, because new theories build gradually on the successes of old ones. Similarly, children's improvement in performance on false belief tests might be explained as a process by which rules are discovered and discarded as better ones come along. But anyone advocating this position needs to formulate an account of the contents of these supposed rules along with evidence that their adoption and replacement at various times would produce the required gradual change in performance levels. Besides, the interpretation offered by Wellman *et al.* has been criticized by Scholl and Leslie (2002) who point out that the meta-analysis only shows that some important change happens at age 4, but does not inform us what this change actually is. Scholl and Leslie maintain that because the results of the meta-analysis are compatible with a number of different accounts they cannot be used as evidence in support of one particular account. Moreover, recent empirical work (Yazdi, German, Defeyter, & Siegal, 2006) on manipulations that make the false belief location more salient (such as asking 'where will Maxi look first') showed an increase in the performance of younger children on the task, contrary to the claims in Wellman *et al.*'s meta-analysis. The finding that the performance of younger children can be improved is incompatible with the conceptual change theory.

Those who believe children acquire rules for mentalizing (and more generally those who believe in conceptual change) need to explain why children start out giving systematically incorrect judgements in a test of false belief. After all, if children start life without knowledge of any rule, their responses would be unsystematic, and their performance would be around chance level. On acquiring a strategy, it would hardly be surprising if children initially hit upon the wrong one; they might, for example, systematically make a judgment about Maxi's belief based on their own knowledge. Eventually they would discover a correct strategy, perhaps using knowledge of Maxi's impoverished informational access in order to infer his false belief. This suggests a developmental trend that is U-shaped, or perhaps J-shaped: children start out giving right or wrong answers, purely based on chance. They may then acquire a rule, which would make their answers systematic, but not necessarily correct; at the end-point of development, upon acquisition of the appropriate rule, children and adults should give consistently correct answers, but none of the results from Wellman *et al.*'s (2001) meta-analysis support such a prediction (Figure 1). Q1

Why, then, do young children give systematically incorrect answers in a test of false belief? Wellman (1990) offered an explanation within the framework of a theory–theory,

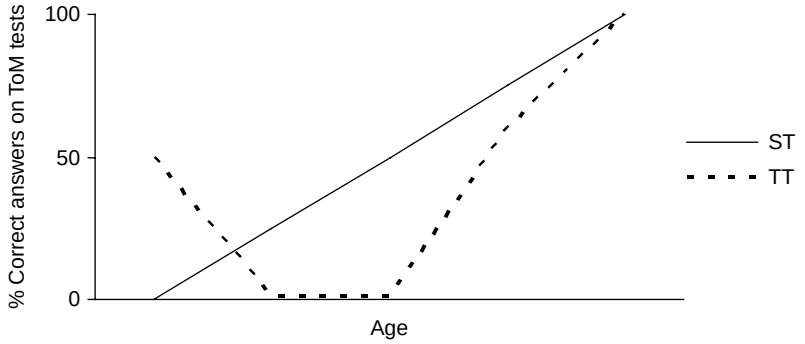

Figure 1. Hypothetical J-shaped trend (dotted line) derived from rule-based accounts: in the absence of a rule children answer at chance level, getting the answer right some of the time. When they acquire a rule, it is likely not to be the right one, so their performance drops to below chance, but upon acquisition of the correct rule, they should reliably answer correctly. Simulation theory predicts a gradual development of performance, with answers systematically incorrect for younger children who fail to quarantine their own knowledge from the simulation process. The actual data (continuous line) follow the predictions of simulation theory (see Wellman *et al.*, 2001).

though his account might not be rule-based. According to Wellman, children begin as 'desire theorists' and then change into 'belief-desire theorists'. In other words, in explaining or predicting behaviour, young children only focus on the protagonist's desires. So, for example, they would predict that Maxi will look for his chocolate in Location B because that is what he would need to do in order to satisfy his desire to get the chocolate. Only older children, according to this account, recognize that Maxi's desire would be thwarted if he held a false belief. The trouble with this account is that it does not explain errors in a deceptive box task. Here, it is fair to presume that children would desire there to be Smarties in the box (in preference to a rather dull pencil). Indeed, we might have expected children's desire to acquire Smarties to help them to acknowledge their false belief. Unfortunately, the data lend no support to this suggestion; indeed, Gopnik and Astington (1988) reported a trend for errors to be even more common in a Smarties deceptive box task than in an unexpected transfer task.

Interim summary: Conceptual change theories cannot account for false belief task performance

According to rule theorists children struggle with the unexpected transfer test of false belief until they acquire the rule which links informational access to consequent states of knowledge. Further, these theorists predict that children may even apply a rule too generally once it is acquired, and there is some evidence in support of this possibility (Sodian & Wimmer, 1987). But, crucially, children demonstrate an understanding that seeing equals knowing well before they can pass a test of false belief (Robinson & Mitchell, 1995), which indicates that acquisition of this rule is not sufficient for false belief understanding.

An additional problem for rule based accounts surrounds children's gradual rather than radical development towards passing false belief tasks. Wellman *et al.*'s meta-analysis indicates that children start out by giving the incorrect response, then give a mixture of incorrect and correct responses before they mostly give correct responses.

This pattern of responses is not easily accounted for by proposing acquisition of a rule, and is perhaps best explained within a simulation account which we turn to in the following section.

Simulation and salience: Accounting for developmental patterns

A rule-based account cannot explain the developmental pattern of children's answers on tests of false belief and it is therefore worth considering an alternative explanation for the performance of children who do not give systematically correct judgements. Within a simulation framework we propose that younger children find it hard to overcome a default to their own mental states. Nevertheless, it might be possible for children to overcome the default if the salience of the other agent's perspective is increased. To overcome a very strong default, presumably young children require especially salient evidence pointing to differences in another person's mental states (or how their own mental states were different in the past). It is thus worth asking whether (a) young children could be supported in making a correct judgment of false belief by manipulating salience and (b) whether older participants (even adults) will systematically confuse their own and another person's belief if the salience of that belief is reduced.

If people do use simulation to predict the mental states and behaviour of others, then it is likely that they will use their own mental states as the default (Harris, 1991). Having such a default offers considerable conceptual economy; instead of having to make a vast number of complicated assumptions about the other person's mental states, you simply assume that their mental functioning is the same as yours, except in certain cases for which you then make allowances (Heal, 1996). This is a reasonable strategy if we assume that other people usually share a common environment, have common concerns and will therefore have the vast majority of their beliefs in common (Fodor, 1992; see also Nickerson, 1999). Creatures that use simulation need, therefore, to take 'no difference in mental state' as the default, from which they shift only when prompted to do so by specific and relevant evidence.[3] If we are asked where Maxi will look for his chocolate, since Maxi is by default likely to be holding a true belief, a simple strategy is to report what we believe to be the current state of reality. Although not a simulation theorist, Fodor (1992) effectively claims something rather similar to Harris, that in order to give a correct judgment, young children must set aside a default to report a salient current reality.

In summary, the idea that currently held beliefs are more salient than beliefs that are not true or current sits comfortably within a simulation account: simulation enables you to exploit your own mental similarity to the target agent in order to avoid having to make a large number of separate assumptions about the target's beliefs and desires (Heal, 1996). Simulation must work, therefore, by making mental adjustments where it is evident that they are needed. An efficient simulation mechanism will evolve only if it has a strong preference for *not* making adjustments.

Thus we have a contrast between an account that makes no reference to salience, according to which children use rules for mentalizing, and an account based on a brand of simulation theory. The latter posits that children's competence develops gradually and can be supported or undermined by the manipulation of associated variables such as salience.

[3] *As Gordon (1986) has argued, simulation is sometimes a useful strategy even when it involves no shift at all; Gordon calls this 'total projection'.*

If an individual had no understanding of minds then presumably they would not reliably give a correct judgment in a test of false belief. But it does not necessarily follow that if an individual gave an incorrect judgment, this implies that they lack understanding of the mind. Admittedly, such an individual is demonstrably not very effective at working out what another person is thinking. But this lack of effectiveness need not necessarily result from lack of understanding that other people have thoughts. As Leslie and Thaiss (1992) pointed out, in order to make a correct judgment of another person's false belief you need to be able to do at least two things: (1) understand the principle that other people hold different and potentially conflicting beliefs and (2) be able to complete a sequence of processing steps without error in order to arrive at a correct diagnosis of what the other person happens to believe. Traditionally, when children have given an incorrect judgment, this has been taken as evidence in support of the view that (1) children do not understand the relevant principle. Another equally plausible explanation, though, is (2) that they make errors in the processing steps.

This problem has arisen because researchers have been over-reliant on children's judgments in a test of false belief for diagnosing competence. And yet it is possible that when children give an incorrect judgment, they nevertheless understand the principle that people hold distinct beliefs. Kikuno, Mitchell, and Ziegler (2007) tackled the problem by measuring how long it took children to answer questions. They reasoned that if children who gave an incorrect judgment lacked a concept of belief, as argued by advocates of the view that mentalizing is based on a theory (e.g. Gopnik, 1993; Perner, 1991), then at least two things should follow: (1) children should treat a question about another person's belief as if it were a question about the physical state of the world (Wimmer & Hartl, 1991) and therefore should answer a belief question as quickly as they answer a question about a matter of fact and (2) children who give a wrong judgment should answer more quickly than children who give a correct judgment, because those who give a wrong judgment supposedly reason merely about the state of the world and do not engage in the more complicated reasoning of what someone thinks about the state of the world. Contrary to these predictions, young children took longer to respond to a question about belief than to a question about a matter of fact and, importantly, they took just as long to respond when they gave an incorrect judgment as when they gave a correct judgment. That is, children apparently take longer to work out what another person is thinking than they take to comment on a matter of fact, and that is true whether they make a correct or incorrect inference of belief. Clearly, then, young children treat questions about belief differently than they treat questions about matters of fact, a finding that is hard to explain for advocates of the theory of mind position.

If young children have a basic competence, even though performance limitations in a standard test of false belief prevent them from expressing this, then we might expect to find signs to this effect in tasks that impose different and fewer demands on performance. Onishi and Bailargeon (2005) found that children as young as 15 months demonstrate false belief understanding in a task utilizing a preferential looking method which removes the need for a verbal component of the false belief task. The infants watched an actor hide a toy in one of two locations. Subsequently a change occurred which meant the actor held a true or false belief about the location, and the critical question is whether infants would expect the actor to search for the toy based on her belief about the location. If the infants expect the actor to search based on her belief they should look longer when that expectation was violated, irrespective of whether the

actor holds a true or false belief. Onishi and Bailargeon found that infants looked longer when the search was not based on the belief the actor had of the toy's location, indicating that this behaviour violated the expectation they had. This seems to indicate that infants understand that people have beliefs.

Further evidence in support of the view that young children have some understanding that people hold beliefs was reported by Clements and Perner (1994; Garnham & Perner, 2001). Children aged around 3 years who did not take into account a protagonist's belief when predicting his search nevertheless revealed sensitivity to the protagonist's belief in their eye-movements. Specifically, when the protagonist held a false belief, participants spent a larger proportion of time looking to the location that actually contained the sought object; when the protagonist held a true belief, in contrast, looking was largely confined to the location that contained the object. Evidently, children were sensitive to the state of the protagonist's belief. However, this apparently was not sufficient for them to give a correct verbal or pointing judgment. Why?

According to Clements and Perner, the children had implicit understanding (revealed by eye-movements) but lacked explicit understanding (incorrect verbal or pointing judgments). In our view, the eye-movements in the false belief condition reveal two things: (1) children were sensitive to the difference between true and false belief conditions and (2) in the false belief condition children spent a fair amount of time looking at the 'false belief' location but they also spent some time looking at the 'true belief location' perhaps suggesting that they were deciding which location to choose (looking from one location to the other). It so happens that the young children sometimes made an incorrect choice (indicated by the verbal judgment or pointing), perhaps because the salience of the incorrect location led to bias. In other words, we suggest that the young children understood the principle that other people hold beliefs, but were prone to error when performing the task of working out what that belief actually was (Kikuno *et al.*, 2007). Moreover, we cannot see that any further explanatory value is conferred by asserting that one kind of behaviour is implicit (eye-movements) and the other is explicit (verbal judgment or pointing). In our view, it is not that the two kinds of behaviour reveal the same thing but on different levels (implicit vs. explicit) but that the two kinds of behaviour tap into different things (eye-movements reveal sensitivity to beliefs; verbal or pointing judgments reveal the consequences of bias that affects performance).

How do people access their own mental states?

An often-heard objection to the simulation account is that it presupposes that children understand their own mental states, whereas the evidence suggests that this is not so. If children understand other minds by running simulations based on their own mental processes, they must, it is argued, be able to access these processes. But Gopnik and Astington (1988) found that children aged about 3 years were unable to acknowledge their own prior false beliefs. Children were presented with a deceptive box test of false belief, in which they were shown a Smarties tube. Initially, children guessed that it contained Smarties, whereupon the experimenter opened the lid to reveal that it contained a pencil. The experimenter returned the pencil, closed the lid and then asked what the child had thought was inside when he or she first saw the tube. Children aged about 3 years wrongly reported the current content (pencil), and many also failed to acknowledge Maxi's false belief in an unexpected transfer test. Failing to acknowledge one's own false belief proved to be at least as difficult as acknowledging

another person's. The conclusion is that knowledge of our own mental processes therefore cannot be a basis for simulation, since it is assumed that in order to understand that another person could hold a false belief, you would first need to appreciate that you yourself could hold a false belief.

A clever study by Wimmer and Hartl (1991) seemed to reveal the basis of children's difficulty acknowledging their own prior false beliefs; it also lent support to the claim that development involves grasping a rule which constitutes a conceptual leap in sophistication. They devised a 'state change' task that is actually a subtle variation on the deceptive box task: the experimenter begins by opening the box to reveal Smarties; the children then see these replaced with pencils. When asked the same question as in a standard deceptive box task, nearly all get the right answer, 'Smarties'.

Wimmer and Hartl (1991) offer this as evidence that these children *lack* understanding of belief: the children do badly on the standard deceptive box task because success requires acknowledging the difference between what was true and what was believed. It was true that the box contained a pencil but children believed it contained Smarties. In contrast, success does not require the children to acknowledge such a difference in state change: it was true that the box contained Smarties and children believed it did.

Wimmer and Hartl conclude that 3-year old are constrained to report reality (in this case, the prior state) specifically because they lacked the capacity for inferring a belief. This suggestion seems much more plausible than the idea that children could not remember what they thought, or that they misunderstood the test question. These memory and communication explanations, if correct, should also apply to the state change task, and therefore we might predict errors at the same level as in a deceptive box task – a prediction that gained no support. Wimmer and Hartl neatly explain how children gave a correct judgment in state change and an incorrect judgment in the deceptive box task; according to them, a conceptual deficit (or lacking the required processing rule – although it is unclear precisely what this rule might be) constrains children to interpret 'What did you think was inside . . .?' as, 'What was inside . . .?' Importantly for Wimmer and Hartl's account, this explains why children's errors in a deceptive box task are *systematically* incorrect and not random, without needing to invoke salience of the current belief as an explanation.

How should a simulation theorist respond to this? State change differs from a standard deceptive box task not only in that the initial belief is true – the aspect seized on by Wimmer and Hartl – but in that the initial belief is made memorable by the visual presentation of the Smarties. It is this, we claim, that explains why the children tend to get the right answer in state change while failing in the standard deceptive box task. They get the right answer in this case because seeing a physical token associated with their prior belief is highly salient and therefore eminently retrievable, which then serves as a mental cue for correctly judging what they initially thought. So on our account, when children answer the question in state change, they are reporting a prior *belief*, not the previous state of the world. Ironically, then, we are suggesting that state change promoted correct judgements by elevating the salience of the initial belief by virtue of its having a physical counterpart.

An obstacle standing in the way of this explanation is that Wimmer and Hartl's (1991) experiment does not allow us to distinguish the effects of a belief's being true from the effects of it having a certain sort of salience. Saltmarsh and colleagues devised an experiment which does this (Saltmarsh & Mitchell, 1998; Saltmarsh, Mitchell, & Robinson, 1995). A box is opened to reveal an atypical content (Smarties tube

containing a key), which the experimenter conspicuously exchanges for another atypical content (pencil) as the child participant watches. What will children say when asked, 'When I first showed you this box, what did you think was inside?' In line with the explanation of the result of their own experiment, Wimmer and Hartl would predict that young children would respond with 'key', since this is the answer you get by considering what was true. This did not happen. Rather, the most common response was to report the current content (pencil), not the first content. Also, young children responded differently depending on whether or not the test question included the word *think*. With its inclusion, they tended to report the current content; otherwise they correctly reported the first content (key). This result is explicable on the assumptions that: (1) the children were genuinely sensitive to the distinction between what is true and what is believed and (2) that their capacity correctly to report a belief is affected by the salience of that belief (in this case, the current belief having greater salience).

In the study by Saltmarsh *et al.* (1995), as in a standard deceptive box procedure, the child's initial belief is not supported by a physical token and therefore is not salient; perhaps this is why children reported the more physically salient current content of the box when asked the standard question that included the word *think*: children would have seen that the current content had a physical embodiment and their belief based around this needed to be set aside in order to give a correct judgment. In a standard state change, the initial belief has a physical embodiment, allowing children to set aside their current belief and so give a correct judgment of belief for the right reason (*pace*, Wimmer & Hartl, 1991).

Another explanation for the results in state change is that children gave a correct judgment because their expectation of the content was confirmed. This differs from the physical salience theory, because it says that correct judgements would be confined to circumstances where expectations are confirmed. The physical salience argument, in contrast, says that salience is sufficient to help children judge correctly, even if their expectation of the box's content is not confirmed.

Results reported by Mitchell and Lacohee (1991; also, see Freeman & Lacohee, 1995) help to clarify matters. In their task, children selected a photo from an array to represent what they thought was inside a Smarties tube. After children had selected the Smarties photo, they posted it in a post box where it remained out of sight. Children had thus supported their own initial belief with a tangible and salient token. Subsequently, the experimenter revealed the true content as a pencil and the task proceeded as in a standard deceptive box task. Children were more likely to give a correct judgment than in a deceptive box task that did not involve posting a picture of Smarties. In this experiment, the children's initial belief was associated with a physical token (the posted picture) even though the belief was subsequently disconfirmed: the box had contained a pencil all along. Follow-up studies, in which children saw the typical content of the box, as in state change, but which nevertheless concerned the current false belief of another person, demonstrated robust improvement in children's judgements (Saltmarsh & Mitchell, 1998; Saltmarsh *et al.*, 1995).

The findings reported above do not show merely that children can acknowledge false belief at an age younger than demonstrated hitherto (cf. Wellman *et al.*, 2001). They show at least two other important things. First, the most promising explanation for systematically incorrect judgements within a non-salience framework (Wimmer & Hartl, 1991) has been refuted. Second, the findings lead to a new way of thinking about children's early performance in handling beliefs. This moves us from the question of

whether children can or cannot acknowledge belief to the question of how salience affects their judgment; in some cases it could lead children to judge incorrectly.

A critic might argue that the salience hypothesis makes predictions identical to a hypothesis of conceptual change that makes no reference to salience, and hence that the notion of salience has no explanatory value. In fact the two hypotheses differ in the way they explain errors in a test of false belief. One who posits conceptual change as a sufficient explanation would say that errors occur where children lack a concept of belief and fall back on reporting what they themselves believe to be true. In contrast, the salience hypothesis says that younger children find it difficult to disengage from their own current beliefs when asked to consider the beliefs of another or their own earlier beliefs. With increasing age they are increasingly able to do this as they become better able to set aside their current belief. In other words, we assume that salience continues to exert an influence on mentalizing, but that its effects are nullified by an improving ability to set aside one's default to current reality. Crucially, the salience hypothesis therefore allows the possibility that older participants make systematic errors in subtle mentalizing tasks where it might not be so obvious that they need to set aside their current belief. This would be difficult to explain by positing conceptual change: it would be difficult to explain the *systematic* tendency to report one's own belief when asked about another person's belief without making reference to salience.

How should a simulationist explain young children's difficulty acknowledging their own false beliefs? Opponents of simulation accounts cite children's difficulty acknowledging their own prior false belief as a sign that the mind is not accessible to itself (Gopnik, 1993). If one's own mental states are inaccessible, the argument goes, it is difficult to see how they could be used in simulations of what other people think, because simulation theory says we use our own mental processes to model, and hence to gain information about, the mental states of others. Two things need to be said in response. First, irrespective of the accessibility of our prior belief (the focus of Gopnik's thesis), we might still have access to *current* mental states, and it might be this particular access that is vital for simulation (Goldman, 1993; Harris, 1993). Second, a process of simulation might be required to work out one's own prior beliefs, just as it is required to work out those belonging to other people.

How would simulation help us work out what we used to believe? For current belief there is a simple method based on the idea that, as the philosopher G. E. Moore noted, you cannot coherently assert P and also deny that you believe P. In order to figure out whether I believe it is raining now I do not need to think about my beliefs at all; I need only think about whether it is raining. Concluding that it is, I can immediately assert 'I believe that it is raining'. This has been dubbed an 'ascent routine' (Evans, 1982; Gordon, 1996).

However, using an ascent routine is not sufficient in itself in order to judge what I used to believe. To achieve that, an ascent routine needs to be combined with simulation. In simulation, we place ourselves imaginatively in the previous situation where we held the belief in question. This does not mean that we look inwards, into a mental store of previously held beliefs. Rather we look outwards, taking our earlier perspective on the world (Heal, 1986). From that imagined position, it may seem to us that P is true – in the world imagined as it used to be, it is raining – and we can then apply an ascent routine to conclude 'I believe P'. Moving out of imaginative mode, we may then conclude that we believed P at the previous moment in time. This gives a simulation-based attribution of belief without introspection (Gordon, 1995). To make a judgment about another person's belief, we use a similar process, taking on,

in imagination, the other's perspective on the world. In both cases the resulting attribution might partly depend on theory: were I to believe that the subject is psychologically very different from me, I might be less willing to use my simulation as a basis for a conclusion about what he or she believes.

In this way we can explain why people are better at attributing beliefs to their current selves than they are in attributing beliefs to their past selves or to others, without invoking introspection and within a simulation framework. The difference between attribution to one's current self and these other attributions is that while all require the use of an ascent routine, attribution to current self does not require the use of simulation.

In summary, young children's difficulty with their own prior false beliefs poses no problem for the view that they work out the contents of beliefs by a process of simulation: that process can be used for working out your own prior belief as well as for working out other people's. Factors that interfere with the process can then be expected to have the same impact, whether you are working out your own prior belief or another person's.

Adults' difficulty with false belief

Results reported by Mitchell, Robinson, Isaacs, and Nye (1996) suggest that salience features in whether or not adults make systematic errors in estimating another person's belief. In their study, a protagonist, Kevin, saw juice inside a jug; later, Rebecca told him that it contained milk. Adult participants then judged what Kevin believed – would he believe what he saw in preference to what he was told? Participants might be swayed by the thought that seeing is more reliable than testimony, in which case they would judge that Kevin believes there was juice in the jug. Or they might be swayed by the information which seems most up-to-date, in which case they would judge that he believes it contains milk. Neither solution logically takes precedence over the other.

In a baseline condition, the vast majority placed greater reliance on the visual source of information, judging that Kevin believed there was juice in the jug, in accordance with what he had seen. In a focal experimental condition, however, judgements were radically different. Here the narrator supplied privileged information, stating that in Kevin's absence, and unknown to him, Rebecca had poured out the juice and replaced it with milk, thus implying that her subsequent utterance was true (though Kevin himself only had Rebecca's word to go on). This had an enormous effect on whether or not participants judged that Kevin would believe Rebecca's testimony; in the focal condition many more participants attributed to Kevin the belief that the jug contained milk. Why should that be, given that the baseline and focal conditions did not differ with respect to anything Kevin knew or believed?

The difference between the two conditions seems to be this: in the focal condition, but not in the baseline condition, participants would naturally form a belief about what was in the jug; what they are told about Rebecca's action implies that the jug actually contained milk and it implies that her ensuing utterance is true. The best explanation for participants' different responses in the two conditions is thus that in one but not the other they themselves had a default belief which they were then inclined to attribute to Kevin, not for any rational reason but simply because it was salient; the effect is difficult to explain without making reference to salience.

These results are further supported by a recent study investigating cultural differences in understanding the mind. The study compared individuals from a collectivist subculture, which stresses conformity, reliability and the importance of

the collective, with individuals from an individualistic subculture which stresses the importance of individuality and uniqueness of the individual (Mitchell, Souglidou, Mills, & Ziegler, 2007). Participants judged whether a protagonist would believe a message about an object's location that was different from the location where the protagonist had last seen it. Participants from a collectivist subculture not only scored highly on a measure of trust, relative to those from an Individualistic subculture, they also tended to judge that the protagonist would believe the message, regardless of whether or not they had privileged information about the truth of the message. It would seem that those from a collectivist subculture project their own more trusting disposition on to others when predicting others' behaviour, and those from an individualistic subculture project their own more sceptical disposition. These results lend themselves to an explanation within a simulationist framework, where people assume that the mind of the target works similarly to their own.

A study by Keysar, Lin, and Barr (2003) offers converging evidence from a very different paradigm, showing that adult participants imputed their own knowledge of an object's existence to another participant: they behaved as if an utterance made by the other participant was referring to this object, even though they explicitly acknowledged that the other participant was ignorant about the object. Again, this could be explained by suggesting that the salience of the adult's own knowledge was leading them to confuse this with what the other person knew.

Because one's own beliefs are salient, setting these aside in the task of estimating what another person thinks will require inhibitory control. If inhibitory control is underdeveloped (as in young children) or impaired (as in cases of brain damage), then we might expect errors in belief attribution to be common. In this context, a study involving a patient with damage to the right inferior and middle frontal gyri comes into focus (Samson, Apperly, Kathirgamanathan, & Humphreys, 2005). This patient encountered difficulty in attributing mental states to others as a result of problems with inhibiting his own perspective. Specifically, he was successful when tested on a low-inhibition false belief task (the participant was not aware of the object's true location), but unsuccessful in high-inhibition false belief tasks (the participant was aware of the object's true location). This supports Carlson and Moses (2001) who report a strong and stable link between inhibitory control and false belief task performance in children. At least some of the errors in false belief tasks stem from a failure to set aside your own knowledge of reality.

The idea of the realist bias was taken up and developed by Birch and Bloom (2003, 2007) as the *curse of knowledge* (which is how the bias is now widely known) to explain the pattern of errors made by both children and adults in attributing mental states to themselves and others. The curse of knowledge describes a bias observed in both children and adults: they are influenced by their own knowledge when assessing that of a more naïve person. Birch and Bloom (2003) report a study with 3–5 year old children in a knowledge-attribution task which demonstrated how children's judgment of a more naïve person's knowledge was systematically influenced by their own knowledge in conditions where they were more knowledgeable. Specifically, children were introduced to a puppet who was familiar with one set of toys, but not the other set. Each toy has an object inside it and children should judge that the puppet would know what was inside the toys he was familiar with, but not know what was in the other set of toys. Crucially, Birch, and Bloom manipulated the children's own knowledge by showing them the content of the toys on half of the trials before asking them whether the puppet would know what was inside. As would be predicted by the curse of

knowledge hypothesis, children overestimated the puppet's knowledge of the content of the toys when they themselves knew what was inside. This tendency was particularly strong for the two younger age groups. The older children were much less influenced by their own knowledge. Birch and Bloom (2003) speculate that overcoming the curse of knowledge requires inhibition of one's own knowledge, something the older, but not the younger children managed to achieve.

This account can be accommodated within simulation theory: in taking the perspective of another, particularly someone who is less knowledgeable or ignorant, we default to our own knowledge and this can bias our simulation process. It also ties in with Samson *et al.*'s (2005) and Carlson and Moses' (2001) ideas on inhibition; performance is worse when participants are cursed by their own knowledge, which requires inhibition. When they are ignorant of the true state of affairs, there is no inhibition required and participants perform better on the task.

On simple tasks even older children who have better general processing capacities can overcome the curse of knowledge; however, on more difficult tasks even adults succumb. In a recently published article Birch and Bloom (2007) report the results from an unexpected transfer task with four locations. Having four locations allowed the manipulations of the participants' state of knowledge, so that they could either be aware of the exact location the object had been moved to, or aware that it had been moved, without knowing the exact location. As predicted by the curse of knowledge hypothesis, adults were more likely to judge that the protagonist would look in a specific container when they had been told that this is where the object had been moved to, than when they were ignorant. This is a further demonstration that adults make 'errors' in false belief reasoning and leads Birch and Bloom (2004; also Birch, 2005) to refute the claim that conceptual change is responsible for children's emerging success in false belief reasoning. Instead the data from their adult and child studies support the view of gradual development, where increased inhibitory control helps people to overcome influences of their own knowledge when judging other's knowledge or actions (see Robinson, Rowley, Beck, Carroll, & Apperly, 2006 for a different interpretation).

Evidence in support of simulation based on counterfactual reasoning

If the simulation account of how we process beliefs is right, it should also shed light on closely related abilities. One of them is counterfactual reasoning. It is surprisingly easy to transform an unexpected transfer test of false belief into a test of counterfactual reasoning. In the classic story, Maxi puts chocolate in the blue cupboard and then leaves the scene. His mother subsequently takes the chocolate and grates some into a cake. Absentmindedly, she returns the chocolate not to the blue cupboard, where Maxi left it, but to the red cupboard. The convention is to ask participants to predict where Maxi will look for the chocolate, but a legitimate question can be posed in counterfactual form: if Mum had not made a cake, then where would the chocolate be now?

How might we answer that question? According to the Ramsey test, I assume the antecedent of the statement, try to integrate the antecedent with my existing belief set (which will involve some revision of my belief set, at points where it is inconsistent with, or in probabilistic tension with, the assumption) and then see whether, from within the perspective of that revised belief set, the consequent looks reasonable (Ramsey, 1950; Stalnaker, 1991). In doing this, I am effectively undertaking a simulation: I simulate belief in the antecedent and let my inferential processes run 'off-line' so as to get to the consequent (or not). But here, instead of simulating what someone else

believes, I simulate what would be true in the counterfactual situation. The Ramsey test does not work for all conditionals (Edgington, 1995) but the question about Mum and the chocolates seems a reasonable candidate for its successful application: imagine a world where Mum did not make the cake and then ask, from within that simulation, where is the chocolate? If the simulation works well, then it will yield the answer, 'In the blue cupboard'.

From the point of view of simulation theory, then, there are common processing demands in a false belief and the counterfactual task (see Figure 2); both employ simulation. In false belief, I start by making an assumption, P (I leave after placing the chocolate in the blue cupboard), which forms the basis for my imaginative projection into Maxi's situation; I then imagine returning, thinking Q: the chocolate is still in the blue cupboard. In the counterfactual task I start with the assumption P: Mum did not make the cake. This leads to the conclusion Q: the chocolate is still in the blue cupboard (in the imagined world where Mum did not make the cake).

A study by Riggs, Peterson, Robinson, and Mitchell (1998) supplied data relating to performance on false belief and counterfactual tasks within a single experimental design. Children performed on a series of unexpected transfer tests, some with a question about a protagonist's belief, and some with a question about a counterfactual. There was an impressively strong correlation between performance on each. This seemed not merely to be because the two kinds of task made similar verbal demands. The correlation remained very strong even when verbal ability as measured by an independent test was partialled out.

However, a simulation theorist would not claim that mental state attribution and counterfactual evaluation are one and the same. As Figure 2 shows, while both involve a simulative step, each also involves processing not shared with the other. In the case of the counterfactual, once I have done the simulative step of assuming the antecedent and inferring the consequent from the assumption plus background belief, I have then to use

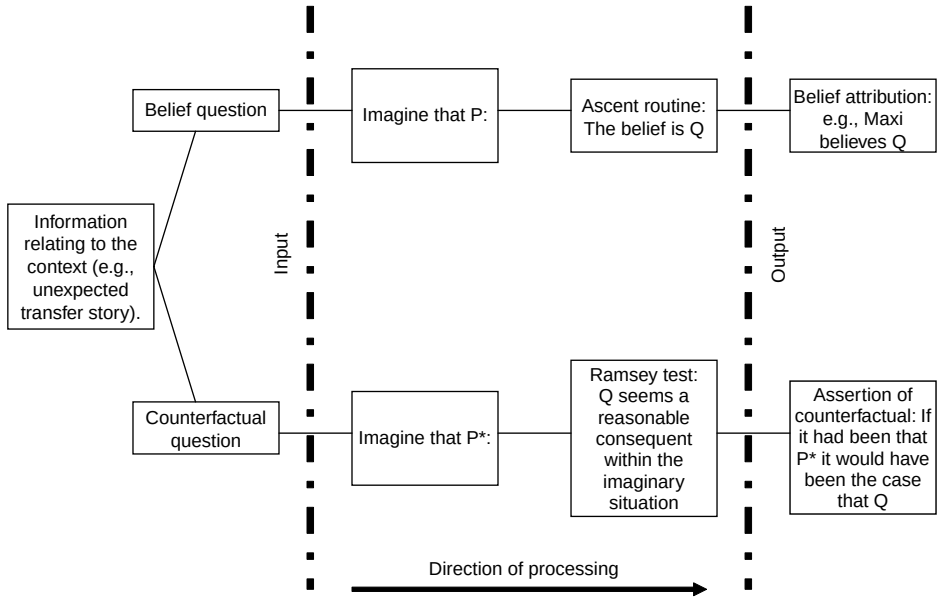

Figure 2. A comparison between how we process beliefs and how we process counterfactuals.

this result to make the judgment that the counterfactual is true. So there is a step here, though a relatively trivial one, from the simulation to a judgment. There is a comparable step to be undertaken in the case of mental-state attribution, but it is done somewhat differently. At the simulative stage, I see the world from the point of view of the protagonist in the story and so I have to make assumptions consistent with the world as she sees it. I have to assume that the chocolate is not moved from the box where I put it. Imagining myself returning to the room, I then find myself, in simulative mode, motivated to look for the chocolate in that box. I can then make a judgment of what the protagonist believes in two further steps, as follows: still in simulative mode, I attribute to myself, via an ascent routine, the belief that the chocolate is in box A; and then, dropping the simulative mode, I infer that this is in fact where the protagonist will believe the chocolate is located. Thus we can see that, while counterfactual evaluation and mental state attribution share a simulative part, they differ as to what is then done with the simulation in order to get to the relevant judgment. And we can say also that, in the case of mental state attribution, this further step is somewhat more complex than in the case of counterfactual evaluation. So we would expect to find situations in which this difference emerges. Are there circumstances under which, or populations for whom, performance on these two tests comes apart, in that the counterfactual evaluation test is performed better than the mental state attribution test?

The answer is yes. Peterson and Bowler (2000) found a mismatch between level of performance in a test of false-belief and ability to assess a counterfactual in children with learning disabilities. Specifically, some of the children were adept at counterfactual reasoning, but failed a test of false belief. The participants seemed to have the requisite ability to simulate, given their success on the counterfactual task, and so their errors on the false belief task might be explained as a specific difficulty with the ascent routine. This provides a context for assessing the performance of those normally developing children who found a counterfactual and a false belief test equally difficult. It suggests that the children in this group who do have difficulty actually have problems specifically with the simulative processing that is common to both tasks.

A remaining concern is that children seem able to entertain and even develop contrary-to-fact states of affairs in their pretence from the age of about 18 months (Harris & Kavanaugh, 1993; Leslie & Frith, 1987), yet apparently continue to have difficulty with counterfactuals until they are about 4 years. For example, children of two can respond appropriately to various moves in a game of make believe; if imaginary water is poured over a toy animal, they can say which animal is wet and also which glass is empty (i.e. the one 'poured' from) even though as a matter of fact both glasses are actually empty. It would be hard to explain how the children are responding without the assumption that they are imagining that the toy is wet and imagining that water has been poured from the glass.

As we have made clear already, evaluating a counterfactual can be assisted by an act of simulation: I simulate belief in the antecedent and see whether I can infer the consequent from my temporarily adjusted belief set. But that is not the end of the story; once I have done this simulation I then have to draw the conclusion that the conditional (if antecedent then consequent) is true. Assumption and inference in a pretend context do not require that extra step; in pretence one does not have to step outside the pretence to think about what, on the basis of the pretence, is true of the real world. Perhaps very young children are capable of the simulation step – hence their competence with pretence – but are not good at drawing real world conclusions from pretence – hence their errors in a test of false belief.

Making sense out of confusion

Several features of a rule-based account of children's mentalizing face anomalous data: the evidence for a sharp improvement in performance at age 4 is questionable; evidence suggesting that salience affects performance falls beyond the scope of rule-based accounts; the suggestion of processing specificity for rules dedicated to mentalizing is contradicted by the finding that normally developing children have as much difficulty with counterfactuals as with handling false beliefs; the claim that failure to acknowledge false belief occurs because the child lacks a rule that links information with mental representation is contradicted by the finding that children also have difficulty when asked about their own prior false beliefs and when asked about counterfactuals. Further, there is a plausible simulation-based theory for at least a good deal of counterfactual reasoning. This account lends itself (with modification) to explaining judgements about false belief. On the other hand, children's acquisition of a rule that links information with mental representation is indicated by their apparent over-application of that rule in Sodian and Wimmer's (1987) study (the phenomenon of 'inference neglect').

Perhaps mentalizing involves a mixed strategy, employing both rule-use and simulation, depending on the kind of problem and on the characteristics of the participant. Further evidence, albeit circumstantial, for the use of rules in mentalizing is reported in the study by Mitchell *et al.* (1996) mentioned earlier. Kevin sees juice in a jug, but later Rebecca says to him that the jug contains milk. In a baseline condition adult participants tended to judge that Kevin would believe what he saw (juice) but in another condition, where participants (but not Kevin) knew that Rebecca's utterance was true, a majority judged that Kevin would believe there was milk in the jug, in accordance with Rebecca's utterance. Interestingly, judgements made by children aged 5 and 8 years were immune to additional information indicating that Rebecca's utterance was true. The children tended to judge that Kevin would believe that the jug contained juice, as he saw, irrespective of whether or not they knew that the jug really contained milk. Although children differed from adults, their immunity to privileged information seems robust given that the same was reported in an earlier study by Perner and Davies (1991; also, Mitchell, Robinson, Nye, & Isaacs, 1997).

It is surprising that adults are influenced by their own knowledge when judging what another person believes, while children are immune from the effects of their own knowledge! One explanation is this: children apply a rule that people assign higher priority to direct evidence (e.g. seeing for yourself) than to evidence from testimony when the two are in conflict (Mitchell *et al.*, 1997). Thus, children solve the problem in a rule-based manner. However, just as with Sodian and Wimmer's (1987) task, applying a rule rigidly has drawbacks. There are circumstances where it would be appropriate to believe what you are told in preference to what you have seen, as for example when your seeing occurred some time ago, where your seeing might be unreliable because you are ill or drugged, or where the thing you are seeing is deceptive or illusory, as in magic shows. Perhaps the adult participants in Mitchell *et al.* (1996) had the sense to appreciate this, and moreover to attribute the same appreciation to Kevin the protagonist who was a victim of conflicting information. In other words perhaps adults, unlike children aged 5 and 8 years, considered that applying a mentalizing rule would not be ideal for solving the problem of identifying Kevin's belief; a rule-based procedure would be regarded as too inflexible in this case.

If adults did not use a rule-based approach, then on what basis did they make a judgment about Kevin's belief? They could try to imagine what they would think if they

inhabited Kevin's informationally impoverished world: perhaps they tackled the problem by deploying mental simulation. Because participants were not using a simple rule, however, they would then face an additional challenge. In trying to simulate Kevin's mental state, participants would need to imagine what they would have thought as if they had not been apprised of the fact that the jug contained milk. It is certainly not wrong to judge that Kevin believes the jug contains milk, and in that circumstance perhaps the obvious response is that his belief accords with what the participant believes to be the actual state of reality. This deserves to be called an obvious response if we follow our earlier suggestion that the world as we actually know it is the salient option we have to set aside when performing mental simulation. Ironically then, in this task children are protected from confusing their knowledge with another person's by taking the simple approach of applying a rather inflexible rule. But adults, embarking on the process of mental simulation, become vulnerable to confusion between their own knowledge and that of another person. Such confusion was further demonstrated in Mitchell *et al.*'s (2007) cross-cultural study. Participants from collectivist subcultures who scored high on a measure of trust were more likely to judge that Kevin would believe Rebecca's message than participants from an individualistic subculture who scored low on a measure of trust. The disposition to trust changes the third-person mental state attribution being made; this subtle influence is well accounted for within simulation theory, but not within a rule-based account. That is, if you are a trusting person, as seems to be the case among members of collectivist subcultures, then your simulations of others will assume that others are also trusting – you will assume that they are trusting just as you are trusting.

A study by Keysar *et al.* (2003) also reports a pattern of results relating to biases in adults' mentalizing that is consistent with there being dissociation between rule-use and simulation. Participants behaved as if an utterance made by the other person was referring to an object whose existence was known only to the participant. But when explicitly asked, participants correctly denied that the other person knew of the object's existence. Presumably, the participants easily judged that the other person did not know about the object according to the rule that the other did not have the necessary informational access. However, perhaps when tackling the more challenging task of interpreting the other person's utterance, participants shifted into simulative mode, which allowed their own knowledge of the object's existence to contaminate their simulation of the other person's perspective.

Further evidence for an interplay between simulation and rule-use comes from a study investigating children's perspective-taking in narrative (Ziegler, Mitchell, & Currie, 2005): children aged between 5 and 9 years were presented with short stories centred on different kinds of protagonists. All stories described a movement of a secondary protagonist into the space occupied by the main protagonist, using the deictic terms *come* and *go*. Children showed systematic errors in recall of these deictics when the presentation of the verb was inconsistent with the perspective of the main protagonist. Signs of perspective taking were present even for an object without agency, but were strongest for an animate agent. These results suggest that perspective taking in narrative and language is partly cued by pragmatics (or rules of language) – given that imaginative projection occurred even when there was no protagonist present whose perspective could be adopted – but beyond that the presence of an animate agent provides an anchor, leading to the strongest imaginative projection. Presumably, simulation is facilitated when a protagonist exists whose perspective can be adopted. What we seem to find here then is that children (and adults) engage in perspective-

taking which is partly driven by rules of language and partly driven by an imaginative projection into the space created by narrative and occupied by the protagonist.

The developmental origins of mind reading: From imitation to simulation

We have already proposed that children start out mind-reading by using simulation. We now inquire about the developmental origins of simulation itself. We propose, in line with Meltzoff (2005), that imitation is the precursor of mind-reading, but in contrast to Meltzoff's thesis, and in line with Goldman (2005, 2006), we propose that simulation, not rule-based theorizing, grows out of imitation. We further propose that rule-based theorizing grows out of simulation, not by supplanting it, but by providing alternative short-cuts in certain cases; simulation therefore remains the primary process for mindreading.

Meltzoff and Brooks (2001; also Meltzoff & Moore, 1977, 1989) report that just a few hours after birth neonates display facial imitation. This, they claim, is evidence for neural mapping between observed and executed movements, which allows neonates to succeed in this cross-modal action: they observe the action of another (poking out the tongue), but they cannot know what this observed action feels like. Neonates succeed in reproducing the action, even though they cannot know from experience what their action will look like. Meltzoff (2005) argues that this first person experience allows infants to learn the relation between their own bodily states and mental experiences, creating a map linking their own mind and behaviour. This map can now be used by infants to understand other minds, because they can draw the analogy that others are 'like me'. Meltzoff proposes that when infants see others acting similarly to how they have acted in the past, they project on to them the mental state that goes with that behaviour. Because the other is 'like me' he can be understood to have mental states similar to my own. Further developments are needed from this stage to acquire an understanding of false belief, that is, any situation in which the other is not 'like me' in some way (e.g. their informational access is different). In sum, Meltzoff proposes that children (and adults) can use their own intentional actions as a framework for understanding the intentional acts of others.

What seems not to have been noticed is that Meltzoff's theories and data are amenable to an account in which imitation leads to simulation, though Meltzoff himself does not subscribe to such a view. If others are 'like me' I can use my own mental and emotional apparatus to simulate their behaviour and gain access to their mental states. Much of Meltzoff and colleagues' data can be interpreted in this way.

The way in which we chart development here is consistent with accounts of how artificial intelligence can learn to mind-read. Biever (2007) reports in the *New Scientist* that Leonardo, a robot built at MIT, can pass a false belief task. And he does so by employing a simulation process which has grown out of imitation.[4] Leonardo (built by Cynthia Breazeal, Matt Berlin, and Jesse Gray) uses face, voice and image recognition software to build (or simulate) a 'brain', and he builds a new 'brain' for every new face he sees. Leonardo proceeds on the assumption that this new brain is guided by the same processes as his own, but might not necessarily have access to the same information; in other words, Leonardo takes his own brain as a model for the other brain and through

[4] *Interestingly, Breazeal, Buchsbaum, Gray, Gatenby, and Blumberg (2005) take Meltzoff to be offering a simulation thesis, assuming him to imply that simulation grows out of imitation as a way to social cognition.*

a process of simulation takes their point of view. In this way he can solve the false belief task, because he does not update the other's brain with the new location of the chocolate.

Mirror neurons as evidence for simulation?

About 10 years ago a new class of neurons was discovered in the premotor cortex of macaques (Gallese, Fadiga, Fogassi, & Rizzolatti, 1996; Rizzolatti & Craighero, 2004; Rizzolatti, Fadiga, Gallese, & Fogassi, 1996). These mirror neurons are special because they fire both when the macaque carries out a motor action and when he observes a conspecific or human carrying out the same action. This system is a prime candidate for a neural substrate of imitation. As Goldman (2006) points out, taking the mental stance of the other monkey in this situation does not require the attribution of mental states and it does not generate imitation behaviour, but it could be the starting-point for simulation.

There is a large body of evidence for structural and functional components of the mirror neuron system in humans, which resonates with a wider range of actions and emotions than the comparable system in monkeys. Gallese (2006) reviews the evidence from a number of different studies investigating the mirror neuron system in humans. This investigation is technically more difficult than in monkeys, because work with humans does not allow single cell recording techniques. Gallese (2006), like Goldman (2006), stresses that there is more to social cognition than social metacognition, highlighting the role of low level, automatic resonance processes. Mirror neurons provide a way of directly understanding the actions of others, because the observer's neurons fire as if they were carrying out the action themselves. Gallese (2006) also cites Meltzoff and Brooks' (2001) work on neonatal imitation as evidence that interpersonal relations are established before the infant has developed a sense of self. All these interactions happen at a basic neural level, which precedes any enculturation or linguistic development (see also Onishi & Baillargeon, 2005). According to Gallese, these early relations enable a process of bootstrapping to take place which fosters the development of cognitive and affective development.

Naturally, we cannot claim that mirror neurons are solely responsible for either imitation or simulation; crucially, macaques have mirror neurons but have not been credited with mindreading or imitation abilities (Gallese & Goldman, 1998). However, mirror neurons might be the evolutionary precursor to mindreading abilities, a necessary but not sufficient structure for imitation and simulation.

Adolphs (2003) suggests that the sensori motor system allows us to understand others' emotions by stimulating the bodily state of what it would feel like to experience that emotion. Supporting this argument, Wicker *et al.* (2003) found that the neural activation was the same whether participants felt disgust because they were presented with a disgust-inducing stimulus or whether they observed someone else's face expressing disgust. In these tasks participants were not required to ascribe representational mental states to others or engage in any kind of reasoning. Instead it seems that in observing others we have an understanding of what they are experiencing, indicated by activation of the same neural regions. This understanding is low-level and does not necessarily have to be governed by the same process as higher-level mindreading. We would not like to make an argument for mirror-neurons being the key to simulation and mind-reading (see also Adolphs, 2006; Apperly, 2008). However, these experimental findings give an indication of widespread low-level resonance of intention

and emotion understanding; in other words, activation of neural systems indicate what seems an almost involuntary process of sharing other's emotions and intentions, even if this resonance does not lead to a representational understanding of their minds.

Saxe (2005) has levelled an argument against simulation theory based both on patterns of errors and the supposed link between mirror neurons and simulation. Whilst acknowledging that mirror neurons act on low-level resonances of emotions and actions, Saxe claims that they, and simulation, are not involved in attributing epistemic mental states to others. To support her stance she adopts the 'argument from error' (Nichols & Stich, 2003), implying that the systematic errors made by mind-readers cannot be accounted for within simulation theory, but must instead be caused by the application of (an incorrect) naïve folk psychological theory. For example, Saxe (2005) cites the work on imputing ignorance by Ruffman (1996). Children equate ignorance with getting it wrong and that this is a sign of rule use leading to error. Goldman (2006), however, contends that confusing ignorance (a lack of true belief) with false belief is a *logical* error made by children, which does not reveal competence or incompetence with mental state understanding. Furthermore, it is possible that children focus too much on the ignorance of the other, which leads to the input to the simulation process of an inappropriate pretend state, namely that of 'being wrong' (Goldman, 2006). Specifically, Gordon (2005) argued that simulation can account for children's systematic error of ascribing a *false belief* when they should assign an *either or belief*; he argues that the only way for young children to withhold or inhibit their own knowledge from their vicarious decision-making process is to negate the knowledge or fact. Only later do children learn to simulate ignorance and can arrive at a statement that acknowledges factual indeterminacy. Gordon (2005) argues that because young children are constrained to negate a fact, they will classify ignorance as 'being wrong' and they will thus arrive at the incorrect answer through a process of simulation. Gordon's argument is a neat and parsimonious explanation which ties in well with the realist bias and curse of knowledge accounts.

Saxe (2005) also aims to disprove any suggestion that errors arise from incorrect inputs to the simulation, which she claims is the strongest argument in support of simulation theory. Whilst simulation theory can be defended in some instances by claiming an input error, Saxe (2005) claims that it cannot do so with respect to inference neglect. She reasons that a psychological rule must therefore account for our mind-reading, along with the systematic errors that occur. Even if we agreed with this in preference to Gordon's simulation-based account of the error, it does not necessarily follow that there is no role for simulation in mentalizing, just because children apparently (mis-)use rules in some cases. As we argued previously, the debate should not pivot on whether or not children use rules, but where these rules emerge from developmentally speaking. We suggest that they are secondary to simulation, that they evolve out of simulation and that they do not supplant simulation.

In principle we would expect that finding activation in the same neural areas when ascribing mental states to self and others would favour a simulation account, whilst finding different areas of activation for mentalizing about self and others would favour a theory account. However, Apperly (2008) points out that a paradigm for distinguishing between simulation and theory–theory should meet the following two conditions: firstly, it must clearly distinguish *self* and *other,* so that we can determine whether participants were making judgements about themselves (1st person) or about another person (3rd person) and, secondly, these judgements need to be based on calculations of mental states (beliefs or desires). In reviewing a number of neuroscience articles

which aimed to distinguish between simulation and theory, Apperly says that tasks developed hitherto satisfied one of the conditions at most, but never both: either they did not involve the appropriate kind of self-other distinction or they did not require participants to calculate mental states (e.g. belief or desires). Apperly therefore concludes that neuroscience can advance our understanding about many aspects of social cognition, but as yet it has not provided any clarification in the simulation theory debate. Because of the difficulty in ensuring that a participant is processing only first-person mental states or only third-person mental states, and given the difficulty in determining whether a judgment was based on calculating a mental state, Apperly is pessimistic about future neuroscience work being able to inform the debate.

Nevertheless, simulation and theory accounts stand as elaborate and compelling edifices that compete to explain one of the most important and essential aspects of the human condition. Whilst it is difficult to test these theories using the techniques of neuroscience or any other approach, it continues to be the responsibility of researchers to rise to the challenge.

Autism: One route to perspective only?

An account that ascribes a role to mirror neurons in simulation and social cognition would be especially compelling if it offered an account of social difficulties in autism. Autism is a neuro-developmental disorder that is marked by impairments in socialization, communication and imagination. Some have argued that a primary cognitive impairment is responsible for secondary deficits in social and emotional functioning, especially the lack of interpersonal connectedness (e.g. Frith, 2003). Despite that, among the three main cognitive theories of autism there is no clear contender to explain all the deficits (Rajendran & Mitchell, 2007). In addition to the features listed above, autism is characterized by pronounced difficulties in imitation, with a propensity for meaningless echoing of others' speech and actions. The discovery of the mirror neuron system and its proposed function in imitation and simulation gave rise to the mirror neuron hypothesis of autism (Williams, Whiten, Suddendorf, & Perrett, 2001). Deficits in early imitation are well documented in autism (Charman *et al.*, 2000; Rogers, Hepburn, Stackhouse, & Wehner, 2003; Rogers & Pennington, 1991), and as Williams *et al.* (2001) point out, imitation and mind-reading are intimately linked within an account based on simulation theory. Rogers and Pennington (1991) suggested that imitation could play a vital role in the development of socio-cognitive skills. Because imitation is impaired in autism, so this could explain why socio-cognitive abilities are also impaired.

Williams *et al.* (2001) note that imitation is an overt process of acting like another person while simulation is a covert, mental counterpart of that process, of putting yourself in the other's shoes and acting or thinking as if you were them. In other words, imitation shares with simulation the need to set aside your own mental or physical state and focus on that of the other person. According to Williams *et al.* (2001), over 20 empirical studies show severe deficits in imitation in autism, ranging from imitating symbolic gestures, the style of a movement or operation, carrying out a thwarted action with the actor's original intent and simple hand and body movements. Typically developing children successfully imitate all these types of actions and Williams *et al.* (2001) therefore conclude that the impairment of imitation in autism is profound.

If mirror neurons play a vital part in imitation and theory of mind abilities (Gallese, Keysers, & Rizzolatti, 2004; Goldman, 2006), then impaired imitation in autism could point to an impaired mirror neuron system (Williams *et al.*, 2001). Converging evidence

for a dysfunctional mirror neuron system in individuals with autism comes from a range of neurophysiological studies. In an EEG study, Oberman *et al.* (2005) showed that control participants but not individuals with autism had suppression in mu frequency in the sensorimotor area. This suppression is thought to indicate mirror neuron activity, which was not found in the ten high functioning individuals with autism tested on a task of hand movement imitation. Even in tasks where individuals with autism succeed in imitating, this might be achieved by recruiting different neural mechanisms, as revealed in an fMRI study that identified activation of different neural structures in autism than in comparison participants (Dapretto *et al.*, 2006). Again, this is consistent with abnormality in the mirror neuron system in autism accounting for difficulties in imitation.

The seminal study by Baron-Cohen, Leslie, and Frith (1985) raised the possibility that all individuals with autism have difficulty imputing beliefs. This turned out not to be the case (Happe, 1995); indeed, high functioning individuals with autism are able to solve complicated false belief puzzles that even present a challenge to many people who do not have autism (Bowler, 1992). Nonetheless, these individuals with autism still warrant their diagnosis and are debilitated by severe social impairments. Clearly, then, mentalizing adds up to a lot more than solving false belief problems.

In the light of these findings, we are faced with the challenge of explaining some of the proficiency shown in mind-reading tasks, in the absence of an aptitude for social cognition in general. Our hybrid account does just that, by proposing that individuals with autism can find a route to mind-reading which is based on theoretical inference, but cannot deal with a novel social situation by taking the empathic stance (i.e. simulation). Our proposal fits with the supposition that individuals with autism have an impaired mirror-neuron system (Gallese, 2006; Williams *et al.*, 2001), thereby preventing the normal development of imitation, which is well-documented in autism (e.g. Rogers *et al.*, 2003).

Impairment in imitation is linked with impairment in simulation, and our model assumes that a rule-based approach to mentalizing grows out of these more basic processes. Is it possible, though, that some high functioning individuals with autism have aberrant development whereby they belatedly acquire rules for mentalizing without a capacity for simulation? If so, then people with autism might perform well on some social cognition tests, particularly those that are amenable to rule-based solution, but do not have the ability to deal successfully with novel problems that fall outside the scope of application of their mentalizing rules; in these cases, solution would depend on a process of simulation, which would be denied to individuals with autism. In short, we are suggesting that insofar as individuals with autism succeed in certain mentalizing tasks, they might well achieve that success via a different route than individuals without autism. The findings reported by Dapretto *et al.* (2006), cited earlier, are consistent with this possibility in suggesting that people with autism use a different neural strategy to achieve the same outcome (in that case, imitation) as comparison participants.

A test of our suggestion could be conducted by presenting a task based on perspective-taking in narrative (Ziegler *et al.*, 2005); in this paradigm we can identify a level of performance that is based on rule-bound, pragmatic perspective-taking, and a level of performance over and above that which is based on identifying with the protagonist in the narrative. We assume that the latter depends on simulation, something that we propose might be specifically impaired in autism. Our prediction, then, is that individuals with autism will adopt a perspective as far as it is possible to do

so using a rule-based process, but that unlike individuals without autism, they will not have any advantage in perspective-taking when that can be achieved via simulation (activated by identification with the protagonist).

Summing up two routes to perspective: Remaining questions

The suggestion that we can use either rules or simulation for solving mentalizing problems invites various questions. In order to address them we begin by supposing that simulation is primary and that children are therefore constrained to start out with that approach early in development, however elementary the problem. For example, in a simple test of false belief, children begin by using simulation, with the consequence that they are prone to report their own current belief (which is typical among children aged about 3 years). Later they notice regularities that link the correct answers they begin to give (from about the age of 4 years) and the circumstances of the task. Finding by way of simulation that the person believes P, I realize that I could simply have reasoned that, since the person saw just P, he believes P. I can now use what I know about a person's informational access to work out what they think, without using simulation. We might subsequently discover that the rule-based approach has limited application, and we sometimes revert to an approach based on simulation as appropriate.

Having set out our unifying account of the processes underlying mentalizing we now turn to the important issue of presenting testable predictions and sketch a response to potentially sceptical questions that we anticipate. We welcome and invite the challenge to have these questions, and therefore our model of mentalizing, empirically tested.

How does it come about that children can use both simulation and theory?

Young children start out making systematically incorrect judgements on a false belief test because they are doing it by simulation and because their simulations at that point are liable to be seduced by the salience of their own current belief. They subsequently develop an improved ability to set aside their knowledge of current reality, which gives rise to an increase in the prospects of accurate mentalizing in general and in judging correctly in a test of false belief in particular. At this point, an initial condition will start to become familiar, thereby allowing a solution to be reached by applying a simple rule: that people retain their beliefs unless they have access to updating information; without access to updating information, people remain ignorant about the current state of things. For mentalizing problems that fall within the scope of this simple rule, it may not be necessary to simulate the person's mental states.

If children start out using one approach to mentalizing, why should they ever bother with an alternative?

Using rules offers a shortcut, and has added value in conferring protection against systematically reporting their own knowledge.

Why aren't there stable individual differences where some people always use rules and some always use simulation?

Using one or the other is not entirely optional and does not follow an arbitrary developmental sequence. Using rules is subordinate to using simulation, so there could be no typically developing person who began with a rule-based approach (though individuals with autism might only ever use a rule-based approach); using rules grows out of using simulation in typical development (but maybe not in autism). There is an asymmetry, though, in that a person could conceivably tackle mentalizing problems

with simulation and never progress to the more efficient method of using rules for simple problems.

If an individual had both methods of mentalizing at their disposal, how would they decide when to use one rather than the other?

Simulation is the default method, but some problems come marked as especially amenable to rule-based reasoning.

Being able to use rules or simulation implies that participants know at the outset what kind of mentalizing problem it is; how could they have that kind of prescience without already having tackled the problem?

The answer to the previous question applies: participants opt for a rule-based shortcut if they detect characteristics of the problem that are likely to make it amenable to solution by that approach, though there is probably no infallible method here, and participants may switch tactics part-way through.

Conclusion

Our account explains young children's systematic errors in tests of false belief by reference to saliency and the characteristics of simulation. It explains why adults but not children make systemic errors in special kinds of mentalizing tasks. It explains why performance at an early point in development correlates with performance on a test of counterfactual reasoning. It explains why children's judgements on somewhat similar tasks a year or so later seem rule-based: children have moved towards using cognitively economical rules for these simple and familiar problems. Unsurprisingly, their use of rules is not always appropriate, as evidenced by 'inference neglect', though rule-use may have unexpected advantages such as immunity to biases that affect adults, as with Kevin and a jug of juice.

We have suggested that a fruitful way to approach developmental issues in mindreading is with the assumption that people use rules or simulation, depending on the demands of the particular problem at hand. In introducing the hypothesis of a flexible strategy we have examined only some mentalizing tasks, and some aspects of developmental profile. Formulating the idea in detail and in full generality will be a much larger undertaking. Perhaps we have said enough to make the strategy seem worth pursuing.

Acknowledgements

Q3 We thank [insert name] for suggesting that conceptual change theory might make U-shaped predictions about development and for suggesting that a rule-based approach might grow out of a Q3 simulational approach. We thank [insert name] for making comments, relating to counterfactual reasoning, on an earlier draft.

References

Adolphs, R. (2003). Cognitive neuroscience of human social behaviour. *Nature Reviews Neuroscience*, *4*(3), 165–178.

Adolphs, R. (2006). How do we know the minds of others? Domain-specificity, simulation, and enactive social cognition. *Brain Research*, *1079*, 25–35.

Amsterlaw, J., & Wellman, H. M. (2006). Theories of mind in transition: A microgenetic study of the development of false belief understanding. *Journal of Cognition and Development*, 7(2), 139–172.

Apperly, I. A. (2008). Beyond simulation-theory and theory-theory: Why social cognitive neuroscience should use its own concepts to study 'theory of mind'. *Cognition*, *107*(1), 266–283.

Baron-Cohen, S., Leslie, A. M., & Frith, U. (1985). Does the autistic-child have a theory of mind. *Cognition*, *21*(1), 37–46.

Biever, C. (2007). The robots with a sense of self. *New Scientist*, *194*, 30–31.

Birch, S. A. J. (2005). When knowledge is a curse – children's and adults' reasoning about mental states. *Current Directions in Psychological Science*, *14*(1), 25–29.

Birch, S. A. J., & Bloom, P. (2003). Children are cursed: An asymmetric bias in mental-state attribution. *Psychological Science*, *14*(3), 283–286.

Birch, S. A. J., & Bloom, P. (2004). Understanding children's and adults' limitations in mental state reasoning. *Trends in Cognitive Sciences*, *8*(6), 255–260.

Birch, S. A. J., & Bloom, P. (2007). The curse of knowledge in reasoning about false belief. *Psychological Science*, *18*(5), 382–386.

Bowler, D. M. (1992). Theory of mind in Aspergers syndrome. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *33*(5), 877–893.

Breazeal, C., Buchsbaum, D., Gray, J., Gatenby, D., & Blumberg, B. (2005). Learning from and about others: Towards using imitation to bootstrap the social understanding of others by robots. In L. Rocha & F. Almedia e Costa (Eds.), *Artificial life* (Vol. 11/1–2, pp. 111–130). Cambridge, MA: MIT Press.

Carlson, S. M., & Moses, L. J. (2001). Individual differences in inhibitory control and children's theory of mind. *Child Development*, *72*, 1032–1053.

Charman, T., Baron-Cohen, S., Swettenham, J., Baird, G., Cox, A., & Drew, A. (2000). Testing joint attention, imitation, and play as infancy precursors to language and theory of mind. *Cognitive Development*, *15*(4), 481–498.

Clements, W. A., & Perner, J. (1994). Implicit understanding of belief. *Cognitive Development*, *9*(4), 377–395.

Dapretto, M., Davies, M. S., Pfeifer, J. H., Scott, A. A., Sigman, M., Bookheimer, S. Y., *et al.* (2006). Understanding emotions in others: Mirror neuron dysfunction in children with autism spectrum disorders. *Nature Neuroscience*, *9*(1), 28–30.

Edgington, D. (1995). Conditionals and the Ramsey test. *Proceedings of the Aristotelian Society*, *69*, 67–86.

Evans, G. (1982). *The varieties of reference*. Oxford: Oxford University Press.

Flynn, E., O'Malley, C., & Wood, D. (2004). A longitudinal, microgenetic study of the emergence of false belief understanding and inhibition skills. *Developmental Science*, *7*(1), 103–115.

Fodor, J. A. (1992). A theory of the childs theory of mind. *Cognition*, *44*(3), 283–296.

Freeman, N. H., & Lacohee, H. (1995). Making explicit 3-year-olds implicit competence with their own false beliefs. *Cognition*, *56*(1), 31–60.

Gallese, V. (2006). Intentional attunement: A neurophysiological perspective on social cognition and its disruption in autism. *Brain Research*, *1079*, 15–24.

Gallese, V., Fadiga, L., Fogassi, L., & Rizzolatti, G. (1996). Action recognition in the premotor cortex. *Brain*, *119*, 593–609.

Gallese, V., & Goldman, A. I. (1998). Mirror neurons and the simulation theory of mind. *Trends in Cognitive Sciences*, *2*(12), 493–501.

Gallese, V., Keysers, C., & Rizzolatti, G. (2004). A unifying view of the basis of social cognition. *Trends in Cognitive Sciences*, *8*(9), 396–403.

Garnham, W. A., & Perner, J. (2001). Actions really do speak louder than words – but only implicitly: Young children's understanding of false belief in action. *British Journal of Developmental Psychology*, *19*, 413–432.

Goldman, A. I. (1993). Competing accounts of belief-task performance. *Behavioral and Brain Sciences*, *16*, 43–44.

Goldman, A. I. (2005). Imitation, mind reading, and simulation. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science. Volume 2: Imitation, human development, and culture* (pp. 79–93). Cambridge, MA: MIT Press.

Goldman, A. I. (2006). *Simulating minds: The philosophy, psychology, and neuroscience of mindreading*. Oxford: Oxford University Press.

Gopnik, A. (1993). How we know our minds – the illusion of 1st-person knowledge of intentionality. *Behavioral and Brain Sciences*, *16*(1), 1–14.

Gopnik, A., & Astington, J. W. (1988). Children's understanding of representational change, and its relation to the understanding of false belief. *Child Development*, *59*, 26–37.

Gopnik, A., & Graf, P. (1988). Knowing how you know: Young children's ability to identify and remember the sources of their beliefs. *Child Development*, *59*, 1366–1371.

Gopnik, A., & Wellman, H. (1992). Why the child's theory of mind really is a theory. *Mind and Language*, *7*, 145–171.

Gordon, R. M. (1986). Folk psychology as simulation. *Mind and Language*, *1*, 158–171.

Gordon, R. M. (1995). Simulation without introspection from me to you. In M. Davies & T. Stone
Q5 (Eds.), *Mental simulation* (pp. 53–67). Oxford: Blackwell.

Gordon, R. M. (1996). Radical simulation. In P. Carruthers & P. K. Smith (Eds.), *Theories of theories*
Q5 *of mind* (pp. 11–21). Cambridge: Cambridge University Press.

Gordon, R. M. (2005). Simulation and systematic errors in prediction. *Trends in Cognitive Sciences*, *9*(8), 361–362.

Happe, F. G. E. (1995). The role of age and verbal-ability in the theory of mind task – performance of subjects with autism. *Child Development*, *66*(3), 843–855.

Harris, P. L. (1991). The work of the imagination. In A. Whiten (Ed.), *Natural theories of mind* (pp. 283–304). Oxford: Blackwell.

Harris, P. L. (1993). First-person current. *Behavioral and Brain Sciences*, *16*, 48–49.

Harris, P. L., & Kavanaugh, R. D. (1993). Young childrens understanding of pretense. *Monographs*
Q6 *of the Society for Research in Child Development*, *58*(1), R5.

Heal, J. (1986). Replication and functionalism. In J. Butterfield (Ed.), *Language, mind and logic*
Q5 (pp. 135–150). Cambridge: Cambridge University Press.

Heal, J. (1996). Simulation, theory and content. In P. Carruthers & P. K. Smith (Eds.), *Theories of*
Q5 *theories of mind* (pp. 75–89). Cambridge: Cambridge University Press.

Hughes, C., Adlam, A., Happe, F., Jackson, J., Taylor, A., & Caspi, A. (2000). Good test–retest reliability for standard and advanced false belief tasks across a wide range of abilities. *Journal of Child Psychology and Psychiatry*, *41*, 483–490.

Keysar, B., Lin, S. H., & Barr, D. J. (2003). Limits on theory of mind use in adults. *Cognition*, *89*(1), 25–41.

Kikuno, H., Mitchell, P., & Ziegler, F. (2007). How do young children process beliefs about beliefs?: Evidence from response latency. *Mind and Language*, *22*(3), 297–316.

Leslie, A. M., & Frith, U. (1987). Metarepresentation and autism – how not to lose ones marbles. *Cognition*, *27*(3), 291–294.

Leslie, A. M., & Thaiss, L. (1992). Domain specificity in conceptual development – neuropsychological evidence from autism. *Cognition*, *43*(3), 225–251.

Mayes, L., Klin, A., Tercyak, K. P., Cicchetti, D. V., & Cohen, D. J. (1996). Test–retest reliability of false belief tasks. *Journal of Child Psychology and Psychiatry*, *37*, 313–319.

Meltzoff, A. N. (2005). Imitation and other minds: The 'Like Me' hypothesis. In S. Hurley & N. Chater (Eds.), *Perspectives on imitation: From neuroscience to social science. Volume 2: Imitation, human development, and culture* (pp. 55–77). Cambridge, MA: MIT Press.

Meltzoff, A. N., & Brooks, R. (2001). 'Like Me' as a building block for understanding other minds: Bodily acts, attention, and intention. In B. F. Malle, L. J. Moses, & D. A. Baldwin (Eds.), *Intentions and intentionality: Foundations of social cognition* (pp. 171–191). Cambridge, MA: MIT Press.

Meltzoff, A. N., & Moore, M. K. (1977). Imitation of facial and manual gestures by human neonates. *Science*, *198*(4312), 74–78.

Meltzoff, A. N., & Moore, M. K. (1989). Imitation in newborn infants: Exploring the range of gestures imitated gestures. *Child Development*, *54*, 702–709.

Mitchell, P., & Lacohee, H. (1991). Childrens early understanding of false belief. *Cognition*, *39*(2), 107–127.

Mitchell, P., Robinson, E. J., Isaacs, J. E., & Nye, R. M. (1996). Contamination in reasoning about false belief: An instance of realist bias in adults but not children. *Cognition*, *59*(1), 1–21.

Mitchell, P., Robinson, E. J., Nye, R. M., & Isaacs, J. E. (1997). When speech conflicts with seeing: Young children's understanding of informational priority. *Journal of Experimental Child Psychology*, *64*(2), 276–294.

Mitchell, P., Souglidou, M., Mills, L., & Ziegler, F. (2007). Seeing is believing: How participants in different subcultures judge people's credulity. *European Journal of Social Psychology*, *37*, 573–585.

Nichols, S., & Stich, S. P. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Oxford University Press.

Nickerson, R. S. (1999). How we know – and sometimes misjudge – what others know: Imputing one's own knowledge to others. *Psychological Bulletin*, *125*, 737–760.

Oberman, L. M., Hubbard, E. M., McCleery, J. P., Altschuler, E. L., Ramachandran, V. S., & Pineda, J. A. (2005). EEG evidence for mirror neuron dysfunction in autism spectrum disorders. *Cognitive Brain Research*, *24*, 190–198.

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

Perner, J. (1991). *Understanding the representational mind*. Cambridge, MA: MIT Press.

Perner, J. (1995). The many faces of belief: Reflections on Fodor's and the child's theory of mind. *Cognition*, *39*, 51–69.

Perner, J., & Davies, G. (1991). Understanding the mind as an active information processor – do young-children have a copy theory of mind. *Cognition*, *39*(1), 51–69.

Peterson, D. M., & Bowler, D. M. (2000). Counterfactual reasoning and false belief understanding in children with autism. *Autism*, *4*, 391–405.

Plunkett, K., & Marchman, V. (1991). U-shaped learning and frequency-effects in a multilayered perceptron: Implications for child language-acquisition. *Cognition*, *38*, 43–102.

Rai, R., & Mitchell, P. (2006). Children's ability to impute inferentially-based knowledge. *Child Development*, *77*(4), 1081–1093.

Rajendran, G., & Mitchell, P. (2007). Cognitive theories of autism. *Developmental Review*, *27*(2), 224–260.

Ramsey, F. P. (1950). General propositions and causality. In *Foundations of mathematics and other essays*. New York: Routledge.

Riggs, K. J., Peterson, D. M., Robinson, E. J., & Mitchell, P. (1998). Are errors in false belief tasks symptomatic of a broader difficulty with counterfactuality? *Cognitive Development*, *13*(1), 73–90.

Rizzolatti, G., & Craighero, L. (2004). The mirror-neuron system. *Annual Review of Neuroscience*, *27*, 169–192.

Rizzolatti, G., Fadiga, L., Gallese, V., & Fogassi, L. (1996). Premotor cortex and the recognition of motor actions. *Cognitive Brain Research*, *3*(2), 131–141.

Robinson, E. J., Champion, H., & Mitchell, P. (1999). Children's ability to infer utterance veracity from speaker informedness. *Developmental Psychology*, *35*(2), 535–546.

Robinson, E. J., & Mitchell, P. (1995). Masking of childrens early understanding of the representational mind – backwards explanation versus prediction. *Child Development*, *66*(4), 1022–1039.

Robinson, E. J., Rowley, M. G., Beck, S. R., Carroll, D. J., & Apperly, I. A. (2006). Children's sensitivity to their own relative ignorance: Handling of possibilities under epistemic and physical uncertainty. *Child Development*, *77*(6), 1642–1655.

Robinson, E. J., & Whitcombe, E. L. (2003). Children's suggestibility in relation to their understanding about sources of knowledge. *Child Development*, *74*, 48–62.

Rogers, S. J., Hepburn, S. L., Stackhouse, T., & Wehner, E. (2003). Imitation performance in toddlers with autism and those with other developmental disorders. *Journal of Child Psychology and Psychiatry and Allied Disciplines*, *44*, 763–781.

Rogers, S. J., & Pennington, B. F. (1991). A theoretical approach to the deficits in infantile autism. *Development and Psychopathology*, *3*, 137–162.

Ruffman, T. (1996). Do children understand the mind by means of simulation or a theory? Evidence from their understanding of inference. *Mind and Language*, *11*, 388–414.

Rumelhart, D. E., & McClelland, J. L. (1986). On learning past tenses of English verbs. In J. L. McClelland, D. E. Rumelhart, & P. R. Group (Eds.), *Parallel distributed processing: Explorations in the microstructure of cognition* (Vol. 2., pp. 216–271). Cambridge, MA: MIT Press.

Saltmarsh, R., & Mitchell, P. (1998). Young children's difficulty acknowledging false belief: Realism and deception. *Journal of Experimental Child Psychology*, *69*(1), 3–21.

Saltmarsh, R., Mitchell, P., & Robinson, E. J. (1995). Realism and childrens early grasp of mental representation – belief-based judgments in the state change task. *Cognition*, *57*(3), 297–325.

Samson, D., Apperly, I. A., Kathirgamanathan, U., & Humphreys, G. W. (2005). Seeing it my way: A case of a selective deficit in inhibiting self-perspective. *Brain*, *128*, 1102–1111.

Saxe, R. (2005). Against simulation: The argument from error. *Trends in Cognitive Sciences*, *9*(4), 174–179.

Scholl, B. J., & Leslie, A. M. (2002). Minds, modules and meta-analysis. *Child Development*, *72*(3), 696–701.

Sodian, B., & Wimmer, H. (1987). Children's understanding of inference as a source of knowledge. *Child Development*, *58*, 424–433.

Stalnaker, R. (1991). A theory of conditionals. In F. Jackson (Ed.), *Conditionals*. Oxford: Oxford University Press.

Wellman, H. M. (1990). *The child's theory of mind*. Cambridge, MA: MIT Press.

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.

Wicker, B., Keysers, C., Plailly, J., Royet, J. P., Gallese, V., & Rizzolatti, G. (2003). Both of us disgusted in My Insula: The common neural basis of seeing and feeling disgust. *Neuron*, *40*(3), 655–664.

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews*, *25*(4), 287–295.

Wimmer, H., & Gschaider, A. (2001). Children's understanding of belief: Why is it important to understand what happened? In P. Mitchell & K. J. Riggs (Eds.), *Children's reasoning and the mind* (pp. 253–266). Hove: Psychology Press.

Wimmer, H., & Hartl, M. (1991). Against the Cartesian view on mind – young childrens difficulty with own false beliefs. *British Journal of Developmental Psychology*, *9*, 125–138.

Wimmer, H., Hogrefe, G. J., & Perner, J. (1988). Children's understanding of informational access as a source of knowledge. *Child Development*, *59*, 386–396.

Wimmer, H., & Perner, J. (1983). Beliefs about beliefs – representation and constraining function of wrong beliefs in young children's understanding of deception. *Cognition*, *13*(1), 103–128.

Wimmer, H., & Weichbold, V. (1994). Children's theory of mind: Fodor's heuristics examined. *Cognition*, *53*, 45–57.

Yazdi, A. A., German, T. P., Defeyter, M. A., & Siegal, M. (2006). Competence and performance in belief-desire reasoning across two cultures: The truth, the whole truth and nothing, but the truth about false belief? *Cognition*, *100*(2), 343–368.

Ziegler, F., Mitchell, P., & Currie, G. (2005). How does narrative cue children's perspective taking? *Developmental Psychology*, *41*(1), 115–123.

Received 3 October 2007; revised version received 19 May 2008

Author Queries

JOB NUMBER: 415

JOURNAL: BJDP

Q1 We have inserted a citation for Figure 1. Please approve or provide an alternative.

Q2 Reference Frith (2003) has been cited in text but not provided in the list. Please supply reference details or delete the reference citation from the text.

Q3 Please specified the author names to be inserted in this context.

Q4 Please supply the page range for the reference Stalnaker (1991).

Q5 We have inserted the page range for the references, Breazeal *et al.* (2005), Gordon (1995), Gordon (1996), Heal (1986), Heal (1996), and Rumelhart & McClelland (1986). Please check and approve.

Q6 Please check the page range for the reference Harris & Kavanaugh (1993).