The Necessity and Impossibility of Simulation

JOSEF PERNER

Laboratory of Experimental Psychology,
University of Sussex, Brighton, BN1 9QG

IN THIS PAPER I address three points. I highlight some of Martin Davies' clarifications on the meaning of simulation which indicate an emerging consensus about how to distinguish simulation from the theory theory. I add an update on our own empirical evidence against children using simulation to typical theory-of-mind tasks and end by distinguishing two types of simulation: content simulation and attitude simulation, which is to capture another emerging consensus that simulation is in some cases virtually necessary while in other cases contested if not impossible.

1. Clarifications of what simulation is

Martin Davies in his paper helps clear up several ambiguities in the meaning of simulation. I found the following three points particularly helpful:

1. The difference between entertaining in imagination hypotheses about mental states and imaginatively adopting that state (for short: 'imaginative identification').

This distinction helps me understand Gordon's (1992c; in press) insistence that simulation — understood as imaginative identifi-
cation — does not require introspective access. However, discussion of this point also made clear Jane Heal's (in press; this volume) point that in order to put simulation to use more needs to be done. There must be introspective access of some sort when the imaginative identification is over and an answer to the question about the other person's mental state is required in the real world. Davies highlights the need for this step by formulating:

2 Strawson's constraint (referential coherence) which says that if 'believing that \( p \)' is to be a proper predicate then saying about myself 'I believe \( p \)' must have the same truth conditions as somebody else saying about me 'he believes \( p \)'.

This makes clear that Gordon's idea of just pretend-being in a certain state cannot substitute for making a judgement about somebody (including oneself) being in that state.

Davies' paper also helps put into clearer words an unease I (Perner and Howes, 1992, p. 74, n. 1) had felt about Stich and Nichols' (1992) presentation of the 'theory theory' by making:

3 The distinction between transitions amongst representational states whose contents themselves concern mental states of the form 'x believes \( p \)' (theory theory position: reasoning about the mind reasoning about a problem domain) versus transitions amongst representational states whose contents are simply \( p \)' (simulation theory: reasoning about a problem domain, the result of which is then attributed to other person).

My way of making the point was to distinguish the transition between mental states from a mental state representing that transition (Perner, 1991). To my mind this distinction is very important because it relates to the distinction between implicit and explicit knowledge at large and, more specifically, to procedural versus declarative knowledge, i.e., knowing how to reason vs. knowing that one reasons that way. So one can characterize simulation as a technique of making parts of one's implicit know-how explicit so that it becomes explicitly attributable satisfying Strawson's constraint.

The question is whether and when we use this technique. Some consensus seems to be emerging among (some) protagonists from both camps that for certain problems it is the only conceivable method and for other problems it is impossible. Between these two extremes is a wide field of cases where it is a largely empirical issue of whether simulation is used. However, my feeling still is that in the typical
problems investigated in theory-of-mind research, simulation is not used.

2. Firming up the empirical evidence

Unfortunately Davies in his paper cut short the discussion of empirical evidence. Although I agree with him that most of the available evidence is inconclusive, it would be helpful to have a discussion of what — from the point of view of a philosopher who tries to decide between the two positions — would be conclusive evidence against the simulation theory. Most of the existing evidence (cited by Stich and Nichols, 1992) pertains to the fact that people make quite different predictions about how people will act in certain situations from how they themselves act in those situations. As Stich and Nichols (1993) concede, simulation theorists can explain this discrepancy by assuming that people base their simulation on wrong ‘pretend inputs’. The evidence against simulation that Deborrah Howes and I (Perner and Howes, 1992) have contributed to this debate has the great advantage that it is not amenable to this ‘wrong pretence input’ counter argument.

In our experiment a character John tells his friend Mary that he will put the chocolates in one of two locations. In Mary’s absence he puts them, say, on the top shelf. In his absence the chocolates are unexpectedly transferred by a third person to the bottom shelf so that John mistakenly thinks they are still in their old place. Consequently, he and Mary are under the mistaken impression that he knows where the chocolates are. Three questions are asked:

Q1 Where does John think the chocolates are?
Q2 If we ask John: ‘John do you know where the chocolates are?’, what will he say?
Q3 If we ask Mary: ‘Mary, does John know where the chocolates are?’, what will she say?

We found that 5 and 6 year old children had no difficulty answering Q1 correctly but tended to answer both Questions 2 and 3 wrongly by saying that John and Mary would say that John didn’t know where the chocolates were. The answers on Q2 are difficult for the simulation theory to explain, because the correct answers to Q1 attest to the fact that children must have assumed the correct ‘pretend inputs’ for
simulating John. So it remains a puzzle why they then should have problems answering Q2.

The implications of our results have been criticized on theoretical grounds by Bob Gordon (1992c) and on methodological grounds by Paul Harris (1991b).

Gordon has two arguments against our findings.

1 He claims that the assumption that children have introspective access is crucial to our argument. Since his position on simulation does not assume such introspective access, our argument is invalid.

Although we did use the expression ‘introspection’ it seems not critical to our argument. In fact, Davies’ clarification of Gordon’s ‘simulation through imaginative identification’ reinforces that point. The prediction that Q1 and Q2 should be of comparable difficulty follows precisely from the assumption that children ‘imaginatively identify’ with John: ‘I put the chocolates on the shelf and leave for the playground. The experimenter approaches me and asks where the chocolates are. I say, “on the top shelf”. She asks me whether I know where they are and I say, “yes I know”.’ If children manage to employ this technique to answer Q1 then I see no reason why they should not also be in a position to answer Q2 correctly. So why do many of the younger ones answer only Q1 correctly?

2 Gordon suggests that children give wrong ‘John doesn’t know’ answers because they assume that John may have doubts about his subjective feeling of knowing (a point also made by Harris, 1991b). He also argues that this tendency should be stronger in younger children because their own doubts (knowing that John is actually ignorant) seep through into their judgment about how John sees matters.

However, this is implausible because if there were a sophisticated concern about John’s doubts about his knowledge then the age trend should go the other way: it is the older children who should be more, not less concerned in this respect. The younger child’s lack of concern for uncertainty has been documented in many areas. It ranges from their difficulty in understanding probability (Piaget and Inhelder, 1951/1975) to their tendency to judge their own guesses as cases of knowledge (see Perner, 1991, chap. 7, for review).

Our experiment raises another potentially interesting problem for the simulation theory which, unfortunately, has not been addressed by any of its critics or commentators. The problem is why Q3 about Mary’s view on John’s knowledge is not much more difficult for children than Q2 about John’s own view. Mary’s view should be more difficult
because, prima facie, it requires an embedding of imaginative identifications: first with Mary and then, within that identification, with John. I do not want to argue that our data are very conclusive, but I want to raise the general question of whether proponents of the simulation theory think that embedded simulation is necessary for answering questions about one person's view on another person's mental state, and if so, how they think such embedded simulation is supposed to work and to what degree it should affect children's difficulty with such questions.

Paul Harris' reaction to our experiment was primarily methodological. One of his fears was that the younger children may have pragmatically glossed the question 'If we ask John where the chocolates are, what will he say?', as 'If we ask John where the chocolates are, will he know?' And there are other problems with our questions, like the difference in form between Q1 and the other two questions.

To help overcome some of these problems Anna Pearson and I have so far tested 30 children between 4 years and 8 months and 5 years 5 months (same age range as younger half of original study) on the same basic stories but giving children explicit answer alternatives for each question:

Q1 If we ask John: 'John, where are the chocolates?', what will he say? Will he say, 'On the top shelf', or 'On the bottom shelf'?
Q2 If we ask John: 'John do you know where the chocolates are?', what will he say? Will he say, 'Yes, I know', or 'No, I don't know'?
Q3 If we ask Mary: 'Mary, does John know where the chocolates are?', what will she say? Will she say, 'Yes, he knows', or 'No, he doesn't know'?

And the answer alternatives were highlighted with two cards. For instance, for Q1 one card showed John saying (in a speech bubble) 'top' and the other 'bottom'. For Q2 one card showed John saying 'Yes' and the other John scratching his head and saying 'No'. The results are very much the same as in the original study except that children were less than perfect on Q1, but their performance was still much better than on Q2 or Q3 as the following table shows:

<table>
<thead>
<tr>
<th>Question</th>
<th>Percent correct</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q1</td>
<td>70</td>
</tr>
<tr>
<td>Q2</td>
<td>33</td>
</tr>
<tr>
<td>Q3</td>
<td>37</td>
</tr>
</tbody>
</table>
In particular, 13 children answered Q1 correctly and Q2 incorrectly, while only 2 children showed the opposite response pattern (McNemar's $\chi^2 = 8.06$, $p < 0.01$).

So my feeling is confirmed that we do not use simulation to solve problems like these. This however does not mean that we, or children never resort to simulation. In fact, following Jane Heal's (in press; this volume) arguments I think that for some problems it is virtually necessary to proceed by simulation. Yet, in other cases simulation may be in principle impossible. In the remainder of this paper I would therefore like to move away from the traditional question whether simulation is or is not the method for understanding other minds and ask whether we can find criteria for distinguishing the problems where we have to use simulation from those where we do without or where it is even impossible to use.

3. Two types of simulation

Empirical evidence against the simulation theory, as presented above, can persuade us at best that simulation is not used in that particular type of mental attribution problem. In fact, there are other examples where it is very compelling (virtually necessary) that we use simulation so that even staunch supporters of the theory theory (Stich and Nichols, in press) agree. For instance, there is the example used by Harris (1992) of how we predict how another speaker of English will judge the grammaticality of English sentences. Then there are reasoning examples like the one that Stich and Nichols (in press) term 'type-2 Harris simulation'. If we know that Sven believes that all Italians like pasta and that he hears that Maria is Italian then we can predict that he will believe that she likes pasta. This can plausibly (plausibly enough so that Stich and Nichols would not want to rule it out on a priori grounds) be simulated by assuming the pretend input:

All Italians like pasta.
Maria is Italian.

then infer within that pretence

Maria likes pasta.

and finally attribute this to Sven as his belief.

Doing it this way counts as simulation according to Davies' require-
NECESSITY AND IMPOSSIBILITY OF SIMULATION

ment that there must be transitions between mental states that deal with the content domain of the simulated person's mental states, i.e., we transit from mock-believing the premises to mock-believing the conclusion. In principle we could do it differently by representing a transition in Sven's mental states, i.e., if Sven believes these premises then he believes the conclusion. This would be applying a theory of reasoning. But as Heal points out, this may only be possible for such simple textbook examples of reasoning where logical form can be separated from the content and where we have some explicit understanding of which statements can be deduced from which. For real life reasoning this is unlikely. There, we have to engage the full capacity of our own implicit reasoning capacity without knowing which precise logical rules we are using (if any).

Stich and Nichols (in press) see the distinctive mark of these plausible, Harris-type cases of simulation as requiring only simulation of other person's belief but not of behaviour. Heal (in press) thinks that a major stumbling block for simulation is having to start from external situations. I think that both these characterizations of the limitations of simulation are not quite right, as the example of predicting one's opponent's move in chess illustrates. As a chess master once remarked, if one plays chess one should forget about the opponent and be concerned with just the relative positions on the board. That is, one can, in the fashion required by simulation, put one's own mind to the description of the external situation (current position) and reason through to the move that the other one should take (and if he reasons the same way I do, will be likely to take). So, here is a very plausible case of simulation which starts from an external situation and includes a behavioral prescription.

As a first stab at an alternative proposal I suggest we distinguish between the following two types of simulation by adopting Davies' terminology:

1 **Content simulation** of another person's mental process involves transitions amongst representational states that have the same content as the simulated person's states. The sequence of contents is usually determined by some rule system (e.g., logic, grammar) that one assumes the other would also apply.

The first thing to notice is that this kind of simulation qualifies under Davies' definition since it involves transition between representational states whose contents concern the subject matter and not the other person's mental states. The other thing to notice is that it can deal
with the cases like the ‘type-2 Harris’ problems mentioned above. For instance, the question about Sven is to figure out the content of his beliefs and not what kind of attitude he holds towards these contents. To determine the contents one employs one’s own mind to draw the logical implications from the premises. The logical implications hold because of the content of Sven’s beliefs, not because these contents are believed rather than, say desired. If Sven desired that all Italians liked pasta and that Maria be Italian, I could still use basically the same simulation to infer that he is likely to desire Maria to like pasta.

This type of simulation, I think, captures quite well Heal’s intuitions about where simulation has its prime role to play: ‘the nub of the simulationist proposal is that in thinking about another’s thought what we do is to take the subject matter of that thought, whether we believe the same or not, and think directly about it’ (Heal, in press). However, it does not feel right to characterize this type of simulation as ‘imaginative identification’ in Gordon’s words. That label has more justification with my second type of simulation:

2 **Attitude simulation** of another person’s mental states involves transition between the same mental states (same attitude + same content) as the simulated person.

This type of simulation is required for those problems which are usually investigated in theory-of-mind tasks. Take for instance the problem of judging whether a person knows the contents of a box depending on whether the person was allowed to look inside it or was prevented from doing so (Wimmer, Hogrefe and Perner, 1988). There is no question about what the content of the person’s knowledge is. The question is what attitude the person has towards a given content, i.e., whether the person *knows* or *does not know* what is in the box. Similarly in our experiment mentioned earlier (Perner and Howes, 1992) the question (Q2) is what attitude John holds towards his own knowledge: does he know he is ignorant or does he think he knows?

However, one might think that the question about where John thinks the chocolates are (Q1, the standard false belief task), could be solved by *content simulation*, since the focus is on *where* he thinks the chocolates are. However — unlike the chess example or the type-2 Harris examples above — the content of John’s belief cannot be derived from the description of story events by following logical rules, rules of grammar or rules like that of chess. To infer where John thinks the chocolates are one would have to *attitude-simulate* his lack of
seeing the transfer to a new location and from that his belief in the old location.

I do not want to claim that *attitude simulation* is impossible, just that it marks those cases that are more contentious than those where *content simulation* is appropriate. One example where we use *attitude simulation* (at least on occasion) is to infer emotions. My well worn example is to imagine myself being followed by a seedy looking character in a dark alley way. This is enough to give me a creepy feeling, to make my heart beat faster and I can feel my pace quicken. As this example demonstrates, *pace* Heal, emotions can be simulated by imagining external situations and, *pace* Stich and Nichols, simulation can include behavioral reactions. Also, the example is a case of *attitude simulation*, since one simulates what emotion the other feels rather than what the other’s emotion is directed at. So it is essential that one goes through the same mental states (attitudes + content) as the simulated person and not just infer consequences from the description of the external situation.

Emotions are, however, different from, in particular, epistemic mental states in that imagined situations elicit real emotions (at least real enough to be classifiable) but do not yield knowledge, beliefs, etc. That observation suggests that simulation of epistemic states may have severe limitations. In fact, it may be impossible to attitude-simulate these states to a sufficient degree of precision so that the simulation yields a clear difference between knowledge, belief, or pretence.

This difficulty can be illustrated with Gordon’s (1986, p. 167) suggestion that,

1. Let’s do a Smith simulation. Ready? *Dewey won the election.*

would be saying the same thing — though less explicitly — as:

2. *Smith believes that* Dewey won the election.

One problem with this suggestion is how we know whether we are simulating that Smith *believes* Dewey won or that Smith is just pretending? One could suggest, that that depends on whether in our simulation we utter a serious statement or a pretend statement. This, however, would be passing the buck, since the critical difference between belief and pretence would be (theory-theoretically) introduced in the decision to utter a sincere or insincere statement.

A similar problem pertains to our John-thinks-he-knows experiment. If I imagine myself in John’s role, putting the chocolates in the
cupboard and then going out to play (and not seeing how the chocolates are transferred) it remains an open question whether I am simulating a belief, knowledge or an imagined assumption (pretence) unless I start my simulation by (theory-theoretically) stipulating that my imagined situations are to be not just imagined but perceived. Again this stipulation entails the critical difference between believing and just imagining. And this critical difference cannot be gained as a result of my simulation but must be theoretically assumed before the simulation can begin. So, when the question is which of these different epistemic states another person entertains then I agree with Heal's contention that simulation cannot proceed from imagined situations.

If it is true that simulation cannot deliver such distinctions, then we have an explanation why experimental results indicate that children do not use simulation in our 'John thinks he knows' experiment. Although one could engage in imaginary identification with story characters (attitude simulation), it would be but idle activity since it would not solve the issue at stake, namely what epistemic attitude Mary or John take towards John's knowledge.

*Note.* The data reported in this paper have been collected with financial assistance from the Economic and Social Research Council research grant R000232886 to N. M. Yuill and J. Perner.
Bibliography


BIBLIOGRAPHY


Perner, J. and Howes, D. 1992: 'He thinks he knows': and more developmental evidence against the simulation (role taking) theory. Mind and Language, 7: 72–86.


