

15/09/2006

Dear Josef, dear Zoltan,

Since I really missed talking to you about implicit knowledge in Sussex I thought I might write a letter describing some ideas I've been working on.

I've focussed on three cases: discrepant performance on false belief tasks—children appear to succeed on a variety of tasks at three or even two years including word learning, deception and violation-of-expectations paradigms, but fail standard false belief tasks—; the 'paradox of early permanence'—four month old infants manifest what appears to be the ability to think about unperceived objects in their looking behaviour but not in their reaching behaviour—; and what I think of as 'causal perception'—infants' looking times show sensitivity to Michotte-like stimuli and also to some basic mechanical principles while their actions and verbal judgements suggest that they lack causal understanding.

In each of these cases we appear to need something like a distinction between implicit and explicit representations of beliefs, objects or causal relations. However, it is not clear whether the distinction we need is the same in every case.

On false beliefs and theory of mind capacities more generally, one candidate explanation for discrepant performance is a distinction between different ways of understanding what beliefs and other attitudes are. To have a belief is, schematically, to bear some relation  $R$  to some entity  $X$ . The  $X$  is normally taken to be a proposition, but one way to distinguish between two kinds of awareness of 'belief' is to suppose that younger children who appear to understand belief really think about it as a relation to an <object, location> pair rather than to a proposition. This is how I understand Ruffman and Perner's response to Leslie. Of course, this can't be the whole explanation of discrepant false belief performance because thinking about 'belief' as a relation between a person and an <object, location> pair is, on the face of it, sufficient

for passing many standard false belief tasks and also for passing false sign tasks (e.g. in Sabbagh et al.'s paradigm, children only need to keep track of the relation between a sign and an <object, location> pair; no propositions are involved). In addition to contrasting understanding belief as a relation to a proposition with understanding 'belief' as a relation to an <object, location> pair, I suggest that we should also consider two ways of understanding the R, that is, the relation someone stands in to an X when X is the content of her belief. On one (pragmatist) way of thinking about this relation, it consists in this: having the belief state leads to actions which are beneficial to the person when X obtains. (By 'X obtains' I mean that if X is a proposition, it must be true; and if X is an <object, location> pair, then the object has to be at the location.) On this view, the content of a belief is taken to be a condition that has to obtain in order for the actions caused by the belief to be successful actions. Clearly this needs refining, but I hope the basic idea is clear enough. On a competing (intellectualist) view about R, having a belief state leads to actions one reason for performing which is X. On this view, the content of a belief is taken to be a condition that determines whether an action is rationally appropriate. I tried to spell out the two ways of thinking about belief and to show that they are genuinely distinct in my 'Awareness of belief' paper (enclosed). I also tried to show how this distinction between two ways of thinking about belief might explain children's apparently discrepant performance on false belief tasks. Very roughly, the idea is that just as philosophers have different ways of understanding what belief is, so might younger and older children.

If these ideas work at all, they can only be part of an explanation of the discrepancy between younger children's success on some false belief tasks and their failure on standard false belief tasks. My idea is that younger children have a pragmatist notion of belief whereas older children have, in addition, an intellectualist concept of belief. However, it seems wrong to suppose that the difference is only a matter of which concepts children have. Surely there might also be a difference in the nature of the reasoning processes involved in the two cases. Very roughly speaking, there might be two 'theory of

mind' mechanisms, one which relies on relatively automatic, possibly modular, inference processes and one which relies on general reasoning.

At this point I think there are two key problems: one is to describe the nature of the processes involved in theory of mind reasoning, the other is to relate the account of the processes to the account of the concepts involved. Suppose (which now seems unlikely to me) that children's successful early 'false belief' performance involves modular cognition. How does the claim that this cognition is modular rather than involving general reasoning processes relate to the claim that this cognition involves a pragmatist belief-like relation between agents and <object, location> pairs?

I attempted to tackle both problems by studying modularity. (I now think modularity is probably the wrong concept for understanding early awareness of mind, but some of the ideas might generalise.) Modularity is generally characterised by reference to a list of properties like information encapsulation and domain specificity; but since modularity is supposed to be an explanatory concept I argued that if there are modules, then there must be a deeper characterisation of modularity which explains why modules characteristically exhibit these properties. One possibility is to characterise modularity by reference to a special kind of computational process, one which is insensitive to what Fodor calls context dependent relations. (A relation between representations is context dependent if whether it holds between two of your representations may depend on which other mental representations you have. For example, the relation ... is adequate evidence for me to accept that ... is a context dependent relation because whether it holds potentially depends on anything you know.) That modular cognition involves a process which is insensitive to context dependent relations can partially explain why modules exhibit domain specificity, informationally encapsulation and restricted access. (It can't explain why modules exhibit neural specificity or innateness, but nor is it clear that modules do exhibit these things.) My idea, then, is that modular cognition and thinking are different

kinds of process, and that we can explain the special features of modular cognition by reference to a special kind of computational process.

Now we have two main ways of contrasting different kinds of awareness of mind: one involves a distinction between concepts, the other a distinction between processes. That is, we have a distinction between pragmatist and intellectualist concepts of belief, and also a distinction between modular cognition and thinking. To find an adequate account of the discrepancy in performance on theory of mind tasks, I suggest we need to combine these distinctions. On the one hand there is a modular capacity to think about relations between agents and <object, location> pairs in terms of a pragmatist notion of belief; and on the other hand, there is a thoughtful capacity to think about relations between agents and propositions in terms of an intellectualist notion of belief. (As I keep saying, I don't think it is actually correct to invoke the notion of modularity in explaining early developing awareness of mind, but it illustrates the idea that an account of the implicit–explicit distinction might need to specify processes.) The problem now is to explain why modular cognition should go with a pragmatist notion of 'belief' whereas thinking goes with an alternative, intellectualist notion of belief. I think this might be achieved by appealing, once again, the claim that modular cognition is not sensitive to context dependent relations.

I'm sorry that this has run on for so long. I was hoping to write about object permanence and causal perception as well, but I fear I have already written too much; anyhow, I have done less work on these cases. In short, I'm hopeful that a contrast between modular and non-modular representations of objects might help explain the 'Paradox of Early Permanence'. And in the case of causal perception it seems to me that discrepant performance might be explained by appeal to the notion of categorical perception—perhaps we perceive some mechanical causings in roughly the way we perceive phonemes and colours. In that case, we may be able to perceive things we can't reason or talk about.

If there is ever another chance to meet up and pick your brains on implicit–explicit distinctions, I'll be very happy.

Best wishes,  
Steve