How to Construct a Minimal Theory of Mind

Stephen A. Butterfill & Ian A. Apperly
<s.butterfill@warwick.ac.uk>

March 30, 2012

Abstract

What could someone represent that would enable her to track, at least within limits, perceptions, beliefs including false beliefs, and other propositional attitudes? An obvious possibility is that she might represent these very propositional attitudes as such. It is sometimes tacitly or explicitly assumed that this is the only possible answer. However we argue that several recent discoveries in developmental, cognitive, and comparative psychology indicate the need for other, less obvious possibilities. Our aim is to meet this need by describing the construction of a minimal theory of mind. Minimal theory of mind is rich enough to explain systematic success on tasks held to be acid tests for theory of mind cognition including many false belief tasks. It is also extensible in ways that can explain a wide range of findings from non-human animals and human infants that are sometimes presented as evidence for full-blown theory of mind cognition. Yet minimal theory of mind does not require representing propositional attitudes, or any other kind of representation, as such. Minimal theory of mind may be what enables those with limited cognitive resources or little conceptual sophistication, such as infants, chimpanzees, scrub-jays and human adults under load, able to track, within limits, facts about perceptions and beliefs.

Keywords: Theory of Mind, False Belief, belief, perception, development, comparative

1. Introduction

What could someone represent that would enable her to track, at least within limits, perceptions, beliefs including false beliefs, and other propositional attitudes? One answer is obvious: she might track these things by virtue of representing them as such, that is, by representing perceptions, beliefs, and other propositional attitudes as such. Our aim in what follows is to identify another, less obvious answer. There is a form of cognition—minimal theory of mind—which does not involve representing propositional attitudes as such but is rich enough to enable systematic success on tasks held to be acid tests for theory of mind cognition including many false belief tasks. As we will explain, this has consequences for interpreting a range of findings concerning infants', adults' and nonhumans' performances on theory of mind tasks. It may also help us to understand what enables those with limited cognitive resources or little conceptual sophistication, such as infants, chimpanzees, scrub-jays and human adults under load, to track, within limits, facts about perceptions and beliefs.

In this section we first defend the theoretical coherence of our question and then explain the findings which motivate facing it.

To many our question may initially appear incomprehensible. Could abilities to track false beliefs (say) really involve anything other than representing false beliefs? To see the possibility of a positive answer it may help to consider a non-mental analogy. What could someone represent that would enable her to track, at least within limits, the toxicity of potential food items? Here the most straightforward answer (she could represent their toxicity) is clearly not the only one. After all, someone might track toxicity by representing odour or by representing visual features associated with putrefaction, say. Suppose Sinéad has no conception of toxins but represents the odours given off by food items and treats those which she thinks smell foul as dangerous to eat, so that she would not normally offer them to friends or family nor conceal them from her competitors. This brings nutritional and competitive benefits thanks to facts about the the toxicity of food. If Sinéad tends to behave in this way because of these benefits, it follows that she has an ability to track, in a limited but useful range of cases, toxicity; and this ability involves representing odour only. Our question, put roughly, is whether false beliefs have something like an odour.

To make the question more precise it is useful to distinguish theory of mind abilities from theory of mind cognition. A *theory of mind ability* is an ability that exists in part because exercising it brings benefits obtaining which depends on exploiting or influencing facts about others' mental states. To illustrate, suppose that Hannah is able to discern whether Isabel's eyes are in view, that Hannah exercises this ability to escape detection while stealing from Isabel, that Hannah's ability exists in part because it benefits her

in this way, and that Hannah's escaping detection depends on exploiting a fact about Isabel's mental states (namely that she cannot usually see Hannah's acts of theft when Hannah doesn't have Isabel's eyes in view). Then Hannah has a theory of mind ability. (This is not supposed to be a plausible, real-world example but only to illustrate what the definition requires.) By an ability to *track* perceptions or beliefs (say), we mean a theory of mind ability which involves exploiting or influencing facts about perceptions or beliefs, respectively. By contrast, *theory of mind cognition* paradigmatically involves ascribing propositional attitudes such as beliefs, desires and intentions to give rationalising causal explanations of action. This distinction is important because the facts about other minds which theory of mind abilities exploit are not necessarily the facts which are represented in theory of mind cognition. To return to the illustration, Hannah is able, within limits, to exploit facts about what others perceive without representing perceptions as such. She has a theory of mind ability while possibly lacking any theory of mind cognition.