

Interacting Mindreaders

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

November 18, 2011

Abstract

1. What is a theory of interpretation?

What grounds ascriptions of thought and action to others? What is the evidential basis for such ascriptions and how does the evidence support the ascriptions? This is the question a theory of interpretation answers.

Answering this question is essential for understanding epistemic aspects of mindreading. For instance, in saying that some principle serves as a heuristic in mindreading we are implicitly relying on a theory of interpretation to identify standards relative to which it is a heuristic.

Note that the question a theory of interpretation answers is a question about justification. It is not about about *how* anyone actually ascribes thoughts or identifies actions. Nor is it (at least not immediately) a question about the mechanisms, processes or representations involved in mindreading. To stress this, Davidson sometimes formulates the issue by asking what someone *could* know that would put them in a position to know another's thoughts and actions (e.g. Davidson 1984b, p. 126).

So a theory of interpretation is just one part of a full account of mindreading. A full account would need to address questions about how mindreaders actually ascribe mental states, about what mental states are, and about the evolution and development of mindreading, among others. Although a theory of interpretation will constrain and be constrained by answers to these questions, none of these questions are directly addressed or answered by a theory of interpretation.

The two most sustained and elaborate attempts to provide a theory of interpretation, Davidson's (1984a) and Dennett's (1987), have ignored the possibility of interaction. The evidential basis they consider and the principles which they identify to link the evidence with ascriptions of

thought and action are available to entirely passive observers. We shall argue that mindreaders who are capable of interacting with the targets of ascriptions are at an advantage. Their ascriptions can be justified by evidence and principles which would be unavailable if they were entirely passive observers.

Being poised to interact with others enables one to know things about their minds which one might not otherwise be in a position to know; or so we aim to show in what follows. Because orthodox theories of interpretation have supposed that being able to interact makes no difference, they have treated the evidential basis for interpretation as more narrow than it truly is. This matters because broadening the evidential basis will not only enable us to give a more comprehensive theory of interpretation; it will also provide tools for explaining the emergence, in evolution or in development, of mindreading.

***Distinguish from ‘second person’, interaction (Gallagher and Hutto?)

To develop the idea that interaction broadens the evidential basis for interpretation we do not need to throw away the leading, most critically examined theories of interpretation due to Davidson and Dennett. An entirely fresh start is unnecessary. Of course we do not mean to assert, positively, that there are no good objections to these theories; perhaps there are. Our point is just that considerations about the role of interaction in theories of interpretation do not give rise to any such objections. A good way to understand how interaction bears on epistemic and normative aspects of mindreading is to extend rather than to reject the leading theories.

Since we shall discuss the role of interaction in theories of interpretation as an extension of leading theories (without any commitment here to the correctness of these theories; any theory of interpretation will have to accommodate considerations about interaction), it will be useful to start with a brief outline of those theories.

2. The leading theory of interpretation

Donald Davidson argues that language is necessary for propositional thought (Davidson 2001, p. 130), and that

References

Davidson, D. (1973 [1984]b). Radical interpretation. In *Inquiries into Truth and Interpretation* (pp. 125–139). Oxford: Oxford University Press.

Davidson, D. (1984a). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.

Davidson, D. (1999/2001). The emergence of thought. In *Subjective, Intersubjective, Objective* (pp. 123–134). Oxford: Oxford University Press.

Dennett, D. (1987). *The Intentional Stance*. Cambridge, Mass.: MIT Press.