

Interacting Mindreaders

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

November 25, 2011

Abstract

1. Davidson's question: What grounds mindreading?

Many people expect to know what those around them think, within limits. If someone were to complain of a co-worker that she can never tell what she thinking, the complaint might be informative, surprising even. And we might reasonably wonder where the fault lies. Maybe the co-worker really is a closed book, but maybe complainant is not really trying to understand her. For humans and perhaps other species too, access to the minds of those nearby, to some of their beliefs and desires, emotions, intentions and knowledge states, is often the norm in this sense: it is lack of access, not access, that is remarkable. And often enough we read others' minds spontaneously without having to stop and think.¹ In this respect we might compare identifying others' thoughts with identifying their masses. And although both can be disguised with the right props and techniques, some degree of revelation is normal.

This mundane fact about minds raises many questions. Some concern mechanism, how one subject is able to discover facts about another's mind. Another set of questions focuses on the evolution of mindreading and the costs and benefits of being a mindreader or a target for mindreading. Our concern here is not with any of these questions. Instead we investigate a more narrowly epistemic question. 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?

To stress that this question is not directly about about *how* anyone actually ascribes thoughts or identifies actions, Davidson (who provided the best

¹ ***fluently — narrative interpretation ***spontaneously — Rubio and Ferguson, Senju et al

developed, most systematic answer to our question) sometimes formulates the question by asking what someone *could* know that would put them in a position to identify another's thoughts and actions (e.g. Davidson 1973, p. 126).

We can clarify this question

The two most sustained and elaborate attempts to answer this question, Davidson's (1984) and Dennett's (1987), do not exploit 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. So on these theories, a purely passive mindreader observing from behind a one-way mirror is on a par with a mindreader who does or could interact with those she seeks to interpret. The two are on a par in this sense: in principle the same evidence could be available to each, and each can exploit the same principles in moving from evidence to ascriptions of thought. We shall argue that mindreaders who are capable of interacting with those they aim to interpret 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 have supposed that being able to interact makes no difference, they have treated the evidential basis for mindreading 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 mindreading; it will also provide tools for explaining the emergence, in evolution or in development, of mindreading.

2. Kinds of mindreading

Davidson's account of what grounds ascriptions of thought is also an account of what grounds ascriptions of meaning. In 'Radical Interpretation' (1973), Davidson suggests that the evidential basis for mindreading (he uses the term 'interpretation') is of this form:

(E) At time t , Ayehsa holds S true and p .

Such evidence confirms or falsifies generalisations of this form:

(G) Ayehsa holds S true if and only if p .

Given the principle that beliefs can be assumed to be true by default, these generalisations are evidence for the conclusion that:

(T) S is true if and only if p .

Knowing this much about what Ayesha's sentences mean enables us to infer her beliefs given that we know which sentences she holds true.

Davidson refined and extended this theory in various ways. By switching from mere correlations between facts and attitudes to causal relations between them, so that the evidence available to the radical interpreter includes facts about *why* subjects hold particular attitudes towards certain sentences, it is possible to draw counterfactual-supporting conclusions about truth conditions.² And by switching the evidential basis from attitudes of holding holding a sentence true to attitudes of preferring the truth of one sentence over another sentence, Davidson was able to show how the evidence he identified grounded inferences concerning what a subject desires as well as what she believes (Davidson 1980).

One striking feature of Davidson's account is that the evidential basis is already intentional. The mindreader (or 'interpreter') starts

3. ***

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

***Mention that being involved in an interaction makes no huge difference to how mindreading works?

To develop the idea that interaction broadens the evidential basis for mindreading we do not need to throw away the leading, most critically examined theories 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; maybe there are grounds for making a fresh start. Our point is just that considerations about the role of interaction in theories of mindreading do not by themselves require a fresh start. 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 mindreading 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.

4. Repetition

Answering this question is essential for understanding epistemic aspects of mindreading. For instance, in saying that some principle serves as a heuristic

² Cf. Davidson 1980, p. 6: 'we are to derive meaning and belief from evidence concerning what causes someone to hold sentences true'.

in mindreading we are implicitly relying on a theory of interpretation to identify standards relative to which it is a heuristic.

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.

5. The leading theory of interpretation

*Not just language – distinguish intentional action from goal-directed in teleological sense; question of interpretation just for how goals are identified. The basic principles proposed by both Davidson and Dennett (or versions of them) apply even to the non-linguistic case.

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

References

- Davidson, D. (1980). Towards a unified theory of meaning and action. *Grazer Philosophische Studien*, 11, 1–12.
- Davidson, D. (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Davidson, D. ([1984] 1973). Radical interpretation. In *Inquiries into Truth and Interpretation* (pp. 125–139). 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.