Detection Theories of Self-Mindreading and the Nature of the Propositional Attitudes

Steve Pearce
Ph.D Candidate,Western University, London, ON, Canada
Spearce7@uwo.ca

Abstract

What underlies our ability to attribute mental states to ourselves? In this paper I will focus on theoretical approaches to self-mindreading that posit a monitoring mechanism which detects the presence of certain kinds of mental states. I will discuss perhaps the most influential version of such a theory, proposed by Shaun Nichols and Stephen Stitch. I will discuss an influential objection to their theory by Alvin Goldman, and then present Goldman's alternative theory. I will argue that Goldman's objection to Nichols and Stich's theory applies to his own as well. Indeed, it applies to any detection theory of self-mindreading. I will diagnose the problem as stemming from an incorrect understanding of the propositional attitudes. Adopting a detection theory of self-mindreading requires that we fundamentally alter our conception of what beliefs and desires are.

1 Introduction

What underlies our ability to attribute mental states to others? Traditionally, two sorts of answers have been given. Some hold that it involves a sort of inference from observed behavior based on a largely implicit theory of the mind. Others hold that it involves simulating the minds of others, making use of the resources of our own minds without the need of a theory.

A special sort of mental state attribution – or *self-mindreading* – is the attribution of mental states to oneself. We routinely form beliefs about our own minds: that we are hungry, that we desire a cup of coffee, and so on. Self-mindreading is theoretically interesting because it seems like we are much better at it. It is a common philosophical idea that we possess a special access to our own mental lives. A theory of self-mindreading must be able to explain this asymmetry.

In this paper I will focus on theoretical approaches to self-mindreading that posit a *monitoring mechanism* which *detects* the presence of certain kinds of mental states. I will discuss perhaps the most influential version of such a theory, proposed by Shaun Nichols and Stephen Stitch. I will discuss an influential objection to their theory by Alvin Goldman, and then present Goldman's alternative theory. I will argue that Goldman's objection to Nichols and Stich's theory applies to his own as well. Indeed, it applies to *any* detection theory of self-mindreading. I will diagnose the problem as stemming from an incorrect understanding of the *propositional attitudes*. Adopting a detection theory of self-mindreading requires that we fundamentally alter our conception of what beliefs and desires are.

2 Stitch and Nichols' Monitoring Mechanism

According to Stitch and Nichols, a theory of self-mindreading needs to explain our capacity for ascribing propositional attitudes to ourselves. The paradigm examples of such mental states are beliefs and desires, but also include fears, hopes and imaginings.

According to the dominant view, propositional attitudes have two components. The first component is their *representational content*. A belief that it is going to rain tomorrow represents that *it is going to rain tomorrow*. The second component is the *attitude type*: e.g. belief, desire, fear and hope. It is this second component that distinguishes a belief that *it is going to rain tomorrow* from a desire that *it is going to rain tomorrow*. Traditionally, attitude types have been given a *functionalist* analysis (Fodor 1975; Loar 1981; Searle 1983). For a mental state to be a desire, for example, is for it to play the "desire-role". This includes, among other things, being apt to generate the appropriate intention when combined with the appropriate belief (Anscombe 1957). It is common to metaphorically refer to the functional roles characteristic of the attitude types as 'boxes'. Thus, a desire that *p* is a mental representation of *p* in the 'Desire Box', for example.

The central explanadum for a theory of theory of self-monitoring, then, is our capacity to quickly, accurately and reliably generate beliefs about our beliefs, desires, and other such mental states. For example, in the appropriate circumstances, when I have a belief that *p*, I am able to form the higher-order belief that *I have a belief that p*. To explain this, Stitch and Nichols propose a 'monitoring mechanism' which "takes the representation p in the Belief Box as input and produces the representation I believe that p as output" (Stich and Nichols 2004, pg. 13).

The operation of this proposed mechanism consists in two steps. First, it detects that there is a belief that *p* present. Second, it takes the content of this belief, and embeds it in a "representational schema" of the form: *I have a belief that* __. Likewise, in the case of the self-ascription of a desire that *p*, the mechanism first detects that there is a desire that *p* present, and then embeds its content in a representational schema of the form: *I have a desire that* __.

And that's it. According to Stitch and Nichols, self-mindreading is achieved by a monitoring mechanism detecting propositional attitudes and embedding their contents in higher-order beliefs. The special epistemic access that we have to our own mental states is explained by the speed, reliability and accuracy of this mechanism.

3 Goldman's Objection

Goldman (2006) raises an important objection to this proposal. He takes issue with the first step of the operation of the proposed mechanism: the detection of the presence of, say a desire that *p*. In order for this to occur, the mechanism must be able to detect that some mental state is a *desire* – that is, it must be able to detect the *attitude type* of the mental state. According to the dominant view, attitude types are functional role. But how, says Goldman, is the monitoring mechanism supposed to detect the functional role of a token mental state? We are never told.

This objection can be strengthened. Functional roles are *dispositional/relational* properties. Part of what it is for a mental state to play the desire-role is for it to be *disposed* to generate the appropriate sort of behavior when combined with the appropriate belief. The functional role of a mental state is a matter of how it is *related* to other mental states, sensory input and behavioral output. But an internal cognitive mechanism can only detect *local* properties of mental states. The attitude type of a mental state cannot be detected by an internal cognitive mechanism for the same reason that we cannot detect whether Jim is taller than Mary just by inspecting the local properties of Jim.

More formally, here is what I will call "Goldman's objection" to Stitch and Nichols' monitoring mechanism theory of self-mindreading:

1. Beliefs about our own propositional attitudes are reliable only if we can reliably detect both their representational content and their attitude type.
2. Attitude types are functional roles, and hence dispositional/relational.
3. Dispositional/relational properties cannot be detected by an internal cognitive mechanism.
4. Stitch and Nichols' theory holds that our beliefs about our own propositional attitudes are generated by an internal cognitive mechanism.
5. Therefore, Stitch and Nichols' theory cannot explain the reliability of our beliefs about our own propositional attitudes.

Goldman suggests that Stitch and Nichols may get around this problem by proposing not a single generic mechanism that detects the attitude type of a mental state, but a distinct mechanism for every attitude type (Goldman 2006, 162). Thus, there would be a "belief-detector" that would detect beliefs, a "desire-detector" that would detect desires, and so on. It is hard to see how this would solve the problem, however. Presumably, the belief-detector is distinct from the desire-detector in that the former is causally sensitive to the functional role of beliefs, while the latter is causally sensitive to the functional role of desires. But the problem remains: an internal cognitive mechanism cannot be causally sensitive to dispositional/relational properties.

4 Self-Mindreading as Quasi-Perceptual

Goldman's alternative theory understands self-mindreading to be quasi-perceptual in nature. On his view, self-mindreading involves the application of *attention* to certain properties of our mental states[1], and the subsequent *classification* of these mental states along a variety of dimensions, including attitude type, representational content, and intensity[2].

According to Goldman, self-mindreading involves a quasi-perceptual process that can be understood as a function from certain input properties of mental states to such varieties of classification. In the case of classification by attitude type, Goldman holds that the input properties are *neurological* (Goldman 2006, 253). On this view, my belief that I am having a *desire*, rather than a *belief,* is the result of the detection of a certain neurological property of my mental states. In order for such a process to result in the reliable classification of attitude type, these neurological properties must be highly – if not perfectly – correlated with the functional role of our mental states.

It is important to note, however, that a similar move is available to Stitch and Nichols. Such a move involves rejecting premise (1) above. Instead of holding that reliable beliefs about attitude type require the *detection* of attitude type, one can hold that it merely requires the detection of a property that covaries with attitude type. The monitoring mechanism need not detect the functional *role* of a mental state, but the neurological properties that *realize* that role.

In order for such an amendment to succeed, neurological properties that are highly – if not perfectly – correlated with the functional roles distinctive of the various attitude types must be

[1] Goldman intends his account to apply not only to the propositional attitudes, but to sensations, emotions and other mental states as well.

[2] We can, for example, believe things to different degrees.

identified. As far as I know, this has not been done. Very little work has been done investigating the neurological differences between the propositional attitudes[3].

One might object that there *must* be such neurological properties, on pain of denying physicalism. Consider, however, that while it is implausible that there is a neurological property perfectly correlated with a mental state representing justice, we do not thereby think that such a mental state has nonphysical features. Representational properties are just not the sort of properties that we should expect to find neural correlates of.

It is not enough that functional roles are neurally realized – no one denies that. Rather, it must be possible for an internal belief detector to be sensitive to all and only those neurological properties that realize the functional role of belief. But practically *any* neurological property can realize that functional role, and the very same neurological property can realize the belief-role in one case and the desire-role in another[4].

So Goldman's alternative theory of self-mindreading fails for the same reason that Stitch and Nichols' does. If attitude types are functional roles, then they cannot be detected by an internal cognitive mechanism: neither directly nor indirectly through the detection of neurological properties.

5 Representational Content and Embedding

According to our best theories of content, representational properties, like functional roles, are dispositional/relational properties. According to causal-informational views (e.g. Fodor 1990), for example, a mental state M represents P just in case M is reliably caused by instances of P. What a mental state represents, then, is (at least partly) a matter of how it is related to the external world.

In light of this, one might argue that an internal cognitive mechanism would likewise be unable to reliably generate beliefs about the *content* of our mental states. The having of a certain content is a dispositional/relational property of a mental state, and internal cognitive mechanisms cannot detect such properties.

To get around this, most detector-based theories of self-mindreading – including both Stitch and Nichols' and Goldman's – hold that the content of the lower-order mental state is *embedded* in the content of the higher-order belief. The monitoring mechanism does not first detect that one has a belief with the content *P*, and then generates the belief that *I am having a belief that P*. Rather, the higher-order belief in some sense *contains* the lower-order mental state, and it is in virtue of this containment that their contents overlap.

Consider a linguistic analogy. Suppose that I wrote "it is raining" on a piece of paper. Suppose that I then appended "I wrote that" in front of the original sentence. I have produced a sentence that correctly identifies the representational content of the sentence that I originally

[3] Some work has been done on investigating the neurological differences between belief and desire *reasoning*. (Liu, Meltzoff, and Wellman 2009) found that event-related brain potential readings were distributed over the mid-posterior scalp during belief and desire judgments and over the right-posterior scalp during belief judgments.

[4] This is the case for representational properties, according to most naturalistic theories of representational content. According to causal-informational views (e.g. Fodor 1990), for example, a mental state M represents P just in case M is reliably caused by instances of P. Such a connection could obtain regardless of the neurological properties of M. And a mental state that represents P in one case could represent Q in another, without changing any neurological properties of M.

wrote. But to do so I did not need to first *detect* what the content of the original sentence was. I could have done this without knowing anything about what the original sentence meant.

Similarly, a self-mindreading mechanism could simply take my belief that *it is raining*, append the content *I believe that* to its content, and produce the belief that *I believe that it is raining*. This produces a belief that correctly identifies the content of my lower-order belief, without the need of first detecting this content.

6 Embedding Attitude Types

Can a similar move be made for attitude types? In the above case, we were able to explain the fact that the content of a belief that *it is raining* reliably leads to higher-order beliefs that *I believe that it is raining* because the content of the former is embedded in the content of the latter. This is to say that the content of the former – *it is raining* – is a *part* of the content of the latter – *I believe that it is raining*.

In order for a similar move to be made for attitude types, we must be able to explain the fact that a belief that *it is raining* reliably leads to a higher-order belief that *I believe that it is raining* and *not* a higher-order belief that *I desire that it is raining*, for example. That is, we must be able to explain the fact that only beliefs reliably lead to higher-order beliefs that represent that *I believe that* ___; only desires reliably lead to higher-order beliefs that represent that *I desire that* ___; and so on.

This cannot be done by appealing to embedding if attitude types are functional roles. This is because *functional roles cannot embed into contents*. According to the dominant view, the property of being a belief is the property of playing the belief-role. But this is distinct from the property of *representing that something is a belief*.

The content of a belief that *it is raining* can be part of the content of a belief that *I believe that it is raining*. But the functional role of a belief that *it is raining* is not part of the content of a belief *that I believe that it is raining*. Functional roles – and hence attitude types, according to the dominant view – cannot embed.

7 Attitude Types are not Functional Roles

The following set of claims is inconsistent:

1. Self-mindreading is accomplished by a detector mechanism.
2. Self-mindreading involves reliable belief about attitude type.
3. Functional roles cannot be detected.
4. Attitude types are functional roles.

Absent an alternative theory of self-mindreading, we should hold onto (1). (2) and (3) seem very plausible. My suggestion, then, is to deny (4). Attitude types are not functional roles. What makes a mental state a belief, rather than a desire, is not a matter of how it is causally related to sensory inputs, behavioral outputs, and other mental states.

What, then, *are* attitude types? What is the difference between belief and desire? A speculative hypothesis that I find compelling is that the difference is *representational* in nature.

According to *attitude representationalism*, beliefs and desires (and other propositional attitudes) are distinguished from on another by their representational content. There is a content distinctive of belief, and a content distinctive of desire. Despite linguistic appearances, a belief that it is raining and a desire that it is raining have distinct contents.

If attitude representationalism is true, then we can straightforwardly explain our capacity to reliably generate beliefs about the attitude type of our mental in the same way we explain our capacity to reliably generate beliefs about the contents of our mental states: by embedding. On this view, an internal cognitive mechanism does not need to first detect that some mental state M is a desire in order to reliably generate a belief that *I am having a desire that* __. Rather, since M being a desire is just a matter of it having a certain content, that content can simply be embedded into the content of the higher-order state. The content of a belief that *it is raining* can be part of the content of a belief that *I believe that it is raining*. Likewise, the "beliefiness" of a belief that *it is raining* can be a part of the content of a belief that *I believe that it is raining*.

This proposal is speculative, and more work needs to be done. In particular, attitude representationalism requires an account of *what* the distinctive contents of the attitude types are. One suggestion comes from moral philosophy. Tenenbaum (2007) argues that desires are "appearances of the good". On this view, a desire that it rain represents, in some way, that *it would be good if it rained*. Thus a desire that it rain is distinguished from a belief that it is raining by its representational content, not its functional role. Self-mindreading of such a desire would simply involve the content *it would be good if it rained* into a higher-order belief.

8 Conclusion

Nichols and Stitch propose a monitoring mechanism theory of self-mindreading. According to this view, the attribution of, say, a desire that it rain to oneself involves detecting the presence of an a desire and embedding its content into the higher-order belief that *I desire that it rain*. Goldman objects to this view on the basis that attitude types are functional roles, and functional roles cannot be detected. He presents an alternative view according to which reliable belief about the attitude type of mental states is accomplished by detecting neurological properties that are highly correlated with their functional role. I pointed out that Stitch and Nichols' can make a similar move, but further argued that such a move fails: there is no reason to think that there are such neurological properties.

I argued that detection-based theories of self-mindreading should abandon the conception of attitude types as functional roles. Instead, we can adopt *attitude representationalism*, according to which attitude types are representational in nature. If we adopt this view, then we can straightforwardly explain our capacity to reliably generate beliefs about the attitude type of our mental states in terms of content embedding.

References

Anscombe, G. E. M. 1957. *Intention*. Harvard University Press.

Fodor, Jerry A. 1975. *The Language of Thought*. Harvard University Press.

———. 1990. *A Theory of Content and Other Essays*. MIT Press.

Goldman, A. 2006. *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*. Oxford University Press.

Liu, David, Andrew N. Meltzoff, and Henry M. Wellman. 2009. “Neural Correlates of Belief- and Desire-Reasoning.” *Child Development* 80 (4): 1163–1171. doi:10.1111/j.1467-8624.2009.01323.x.

Loar, Brian. 1981. *Mind and Meaning*. Cambridge University Press.

Searle, John R. 1983. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press.

Stich, Stephen, and Shaun Nichols. 2004. “Reading One’s Own Mind: Self-Awareness and Developmental Psychology.” In *New Essays in Philosophy of Language and Mind, Canadian Journal of Philosophy, Supplementary Volume 30*. Vol. 34. Supplement. University of Calgary Press.

Tenenbaum, Sergio. 2007. *Appearances of the Good: An Essay on the Nature of Practical Reason*. Cambridge University Press.