A Difficulty for Testing the Inner Sense Theory of Introspection

Michael Roche*†

A common way of testing the inner sense theory of introspection exploits the possibility of damage to inner sense. Such damage is expected to lead to first-personal deficits/impairments of one kind or another. I raise various problems for this way of testing the theory. The main difficulty, I argue, stems from the existence of the method subserving confabulation.

1. Introduction. According to the inner sense theory, we have a special, quasi-perceptual method for detecting our own mental states. This method is contrasted with the indirect, highly interpretive method that we use to attribute mental states to others. Recent defenders of the inner sense theory include David Armstrong (1981), William Lycan (1996), Shaun Nichols and Stephen Stich (2003), and Alvin Goldman (2006).

The inner sense theory is commonly thought to predict the existence of individuals with first-personal deficits/impairments of one kind or another. One can thus test the theory by seeking out such deficits/impairments. My focus in this article is primarily on the first, predictive step of this testing procedure. I argue that various first-personal deficits/impairments that might appear to be predicted by the theory are not so predicted. Moreover, my arguments show that those deficits/impairments that are legitimately predicted by the theory will be quite subtle, and so difficult to detect. There is thus a significant difficulty facing this way of testing the inner sense theory.

*To contact the author, please write to: Department of Philosophy, Illinois Wesleyan University, 128 Center for Liberal Arts, Bloomington, IL 61702; e-mail: mroche300@gmail.com.

†In writing this article, I have benefitted greatly from discussions with Sarah Paul, Lawrence Shapiro, and Danielle Wylie. Thanks also to the audiences at the Southern Society of Philosophy and Psychology (2011), the Society for Philosophy and Psychology (2011), and especially the Philosophy of Science Association (2012), where I presented various versions of this article. In each case, the ensuing discussion was very helpful.

Philosophy of Science, 80 (December 2013) pp. 1019–1030. 0031-8248/2013/8005-0035$10.00
Copyright 2013 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/673967 Published online by Cambridge University Press

2. The Inner Sense Theory: Two Recent Accounts. Nichols and Stich (2003) offer a fairly simple explanation of the seemingly special, direct access that we have to our own mental states. They claim that to explain such access to, for example, our beliefs, all that is needed is a monitoring mechanism that: (i) copies the representations in one's "Belief Box," (ii) embeds the copied representations in an 'I believe that ____' representation schema, and (iii) places the new metarepresentations into one's Belief Box. This approach can be extended to other mental state types by simply adding the appropriate representation schemas: an 'I desire that ___' schema (for desires), an 'I intend that ____' schema (for intentions), an 'I am having a visual experience that ____' schema (for visual experiences), and so on. Nichols and Stich leave open whether there is a single monitoring mechanism equipped with representation schemas for all mental state types, or multiple monitoring mechanisms, each specialized for a certain mental state type.

Goldman (2006) offers an account of inner sense that is quite similar to Nichols and Stich's. There are, however, a few differences worth noting. First, Goldman opts for a single mechanism of inner sense, denying multiple monitoring mechanisms on the grounds of parsimony. Second, unlike Nichols and Stich, Goldman explains how the monitoring mechanism selects the appropriate representation schema within which to embed a copied representation. Notice that Nichols and Stich's boxology is metaphorical; a "box," for them, refers to a functionally characterized set of mental states. For this reason, a monitoring mechanism cannot select the appropriate representation schema by noting the "box" within which a mental state resides. For Goldman, the monitoring mechanism recognizes a mental state's type via its neural properties.

3. Testing the Inner Sense Theory. If there is a mechanism of inner sense, then damage to this mechanism should be possible. Moreover, proponents and opponents of inner sense appear to agree that such damage should bring with it first-personal deficits/impairments of one kind or another. The presence or absence of such deficits has thus been used to test the theory.

3.1. Nichols and Stich's Argument for Inner Sense. I just explained that, for Nichols and Stich, inner sense is constituted by one or more monitoring mechanisms. Importantly, they regard these mechanisms as being distinct from those mechanisms that subserve 'mind reading', that is, our ability to attribute mental states to others. Given this distinctness, Nichols and Stich claim that either could be damaged while the other is not. They write that "on our theory it is possible for one or more of the [monitoring mechanisms] to malfunction, causing a deficit in one or more aspects of first-person mental-state detection, while the [heterogeneous collection of

https://doi.org/10.1086/673967 Published online by Cambridge University Press

mental mechanisms that subserve mind reading] continue to function normally" (2003, 188).[1]

This is the first, predictive step of their argument. They go on to claim that this prediction is borne out, thus confirming the inner sense theory.[2] I will say a bit about this second step in section 7.1.

3.2. Engelbert and Carruthers's Argument against Inner Sense. Mark Engelbert and Peter Carruthers (2010) reject the inner sense theory, taking aim at the two versions of the theory described above. Their central claim is that, on both accounts, inner sense is *dis*unified. Inner sense is constituted by either multiple distinct monitoring mechanisms (one for each mental state type), or a single mechanism with multiple distinct channels (one for each mental state type).[3] Such disunification, they claim, makes possible selective damage, that is, damage to some mechanisms/channels, but not others. They allege that such damage should lead to "people who can self-attribute beliefs but not desires, or who can self-attribute visual experiences but not auditory ones, and so forth" (Engelbert and Carruthers 2010, 246).

This is the first, predictive step of their argument. They go on to claim that this prediction is not borne out, thus disconfirming the inner sense theory.[4]

3.3. The Two Predictions. Making explicit the predictions from these arguments will be useful in what follows:

> **Nichols and Stich's Prediction**. There should be individuals who suffer deficits in first-person mental state detection, but not third-person mental state detection.
>
> **Engelbert and Carruthers's Prediction**. There should be individuals who are capable of self-attributing certain types of mental states, but incapable of self-attributing other types.

1. They also note the possibility of the opposite pattern of dissociation. Because I am concerned here only with arguments for/against the inner sense theory that exploit the possibility of damage to inner sense, I will set aside this direction of their argument.

2. They are primarily concerned with arguing against the 'theory theory of self-awareness', according to which a single method subserves both first-personal and third-personal mental state detection. Their account of inner sense is but one way of denying this theory.

3. For the sake of argument, I would like to grant their claim that a single monitoring mechanism must have distinct channels.

4. Their aim in attacking this theory is to support what Carruthers (2009) calls the 'mindreading is prior model'. For current purposes, this model can be seen as equivalent to the theory theory of self-awareness mentioned in n. 2.

https://doi.org/10.1086/673967 Published online by Cambridge University Press

Notice that while the prediction from Engelbert and Carruthers's argument concerns a specific first-personal deficit, namely, the inability to self-attribute certain kinds of mental states, Nichols and Stich's prediction does not. Their prediction is compatible with various first-personal deficits/impairments, including but not limited to the one described by Engelbert and Carruthers. Other such deficits/impairments will be considered in section 6. Until then, I want to focus solely on the deficit/impairment described by Engelbert and Carruthers.

4. Engelbert and Carruthers's Argument and the Dual Method Theory. Engelbert and Carruthers claim that damage to inner sense should lead to individuals capable of self-attributing certain types of mental states, but not others. Importantly, this claim relies upon the following auxiliary assumption:

> [**Auxiliary Assumption**]. Humans have no method for attributing mental states that (i) can be applied to oneself and (ii) is independent of any alleged mechanism of inner sense.

The sense of independence stated in condition (ii) is important. A method for attributing mental states satisfies this condition so long as it can operate without inner sense. A method's satisfying condition (ii) is thus consistent with inner sense's being ordinarily involved in its operation; it need only be such that it *can* operate without inner sense. If [Auxiliary Assumption] is false, then an individual with a damaged inner sense should be capable of self-attributing mental states via this other method.

The role of [Auxiliary Assumption] in Engelbert and Carruthers's argument is a bit unclear. As opponents of the inner sense theory, they would certainly deny it. But the inner sense theorist is free to deny [Auxiliary Assumption] as well. Indeed, many inner sense theorists, including both Goldman and Nichols and Stich, endorse what is called the 'dual method theory'.[5] Such theorists believe that humans sometimes self-attribute mental states via an indirect, interpretive method of the sort used to attribute mental states to others. If the dual method theorist takes this interpretive method to be independent of inner sense (in the sense explained above), then she will reject [Auxiliary Assumption].

In any case, [Auxiliary Assumption] is false, as I will argue next. Moreover, its falsity in no way jeopardizes the inner sense theory. The inner sense theorist should thus reject [Auxiliary Assumption], thereby undermining Engelbert and Carruthers's argument against that theory.

5. This name comes from Goldman (2006).

https://doi.org/10.1086/673967 Published online by Cambridge University Press

5. An Argument against [Auxiliary Assumption]. The argument in this section is based on the phenomenon of confabulation. A person confabulates when she self-attributes a mental state that is not present, but which, if present, would help to make sense of his or her behavior. Confabulations are the products of an interpretive method that generates beliefs about one's mental life based on evidence concerning one's behavior and situation. Much psychological work has demonstrated the occurrence of confabulation for a variety of mental state types, including intentions (Wegner and Wheatley 1999), judgments (Nisbett and Wilson 1977; Eagly and Chaiken 1993; Gazzaniga 1995), and decisions (Brasil-Neto et al. 1992). These studies, and others, are discussed in Carruthers (2009).

In Wegner and Wheatley (1999), participants and confederates shared control of a mouse connected to a monitor displaying various pictured objects. Participant-confederate pairs took part in multiple trials, each consisting of a period of joint movement and a period of joint stopping. Both participants and confederates wore headphones and were told that they would hear one spoken word per trial. Each spoken word named an object, although only sometimes was that object pictured on the monitor. Participants were told that the words were intended to serve as distractors, while, in fact, they were meant to prime thoughts in the participants about the objects named. Participants were under the false impression that their confederate partners were hearing the same kind of recording, but with different words spoken at different times. In fact, confederates were being instructed by the experimenters to make particular movements at particular times.

In forced stop trials, the confederate would manipulate the mouse so that the cursor would come to a stop on the pictured object named over his partner's headphones. In unforced stop trials, the confederate would not manipulate the mouse. Importantly, in the unforced stop trials, there was no statistically significant difference between where the cursor stopped when the participant's word named an object pictured on the monitor and where it stopped when it did not. This was interpreted as showing that the priming did not cause the participants to form intentions to stop their cursors on the pictured objects named.

Nevertheless, in forced stop trials, participants were on average inclined to (falsely) self-attribute the intention to stop their cursor on the pictured object named over their headphones. Much more importantly, though, the degree of intentionality self-attributed in these cases varied in ways strongly suggesting that the confabulations were produced by an indirect, interpretive method. When the priming word was heard before the (forced) stop, the strength of intentionality self-attributed was inversely proportional to the amount of time between when a participant heard the priming word and when the cursor stopped on the corresponding pictured object. Participants

https://doi.org/10.1086/673967 Published online by Cambridge University Press

were unwilling to (falsely) self-attribute an intention when the priming word was heard one second after the (forced) stop.

Note the reasonableness of the confabulated intentions. In a forced stop trial, the cursor stopped on the pictured object that the participant was primed to be thinking about. In such a case, the interpretation that one intended to stop the cursor on that object is perfectly reasonable. This reasonableness is even greater if the thought immediately preceded the stop. And, indeed, this was reflected in the judgments of the participants. In more ordinary contexts, such interpretations would likely be true.

These points strongly suggest that the self-attributions in the forced stop trials were produced by an indirect, interpretive method sensitive to at least two factors: (i) whether the cursor stopped before or after the word was heard and (ii) the amount of time between when the cursor stopped and when the word was heard. There are numerous studies, like the one just described, that demonstrate confabulation for a variety of mental state types. Taken together, such studies make a strong case for the existence of an indirect, interpretive method for self-attributing mental states. Because such a method is orthogonal to the direct, perception-like method that inner sense is alleged to be, these studies also suggest that this method is independent of inner sense, thus undermining [Auxiliary Assumption]. Finally, note that this conclusion stands regardless of how often humans confabulate.[6]

5.1. The Simulation Theory. I must admit that consistent with what I have argued so far is the possibility that confabulation is subserved by a method that involves inner sense in a way that violates condition (ii). One way of motivating this possibility is to recognize that inner sense plays an important role in some accounts of mind reading. The simulation theory is such an account.[7] But, as I will now explain, there is good reason to think that on this theory, inner sense plays only a dispensable role in mind reading. If so, then condition (ii) is not violated.

The basic idea behind the simulation theory is that one understands what another individual, S, is thinking, how S will act, or why S acted as she did, by "stepping into S's shoes." Consider the following oversimplified example, based on Goldman (2006, 23–53). Suppose you see S behave in some way, X. In order to understand S's (intentional) behavior, you try to imagine what S's mind was like prior to that behavior. Suppose you generate various hypotheses, including the hypothesis that S had a certain belief, Y, and a certain desire, Z. The next step is to test these various hypotheses by running

6. Dual method theorists typically downplay the prevalence of confabulation. Moreover, the confabulation data are typically the reason why the inner sense theorist endorses the dual method theory; see, e.g., Goldman (2006).

7. Not all simulationists are committed to inner sense. See, e.g., Gordon (1995).

https://doi.org/10.1086/673967 Published online by Cambridge University Press

them through your own decision-making processes. Inner sense contributes to mind reading by detecting the outcome of this procedure.

Suppose that inner sense detects that the outcome of testing the Y-Z belief-desire hypothesis is a decision to X. Given that this decision matches S's apparent decision (as evidenced by S's doing X), this is evidence in favor of that particular hypothesis. The final step is to attribute this belief-desire pair to S. I shall call mental state attributions arrived at in this way 'simulation-based attributions'.

As we have just seen, simulation-based attributions are not arrived at purely by simulation. Prior to the simulation, a mind reader must generate various mental hypotheses. Moreover, because there is an indefinite number of mental hypotheses that can make reasonable S's doing X, the generated hypotheses must be narrowed. Mind readers, on this account, must have some way of intelligently narrowing (or framing) the pool of possible hypotheses.

There is thus significant pressure for simulation theorists to posit a component of mind reading that is capable on its own of generating and narrowing reasonable hypotheses about the mental lives of others. Presumably, then, on this theory, if one's inner sense were to become damaged, one would still be capable of attributing mental states to others in a fairly reliable way. All that follows is that these attributions would not be simulation based. For this reason, even if confabulation is subserved by a simulationist method, and even if this method normally involves inner sense, the method likely satisfies condition (ii). Accordingly, the simulation theory does not pose a threat to my argument against [Auxiliary Assumption].

6. Revisiting the Predictive Step of the Two-Step Argument. The result of the previous section is that the method subserving confabulation undermines [Auxiliary Assumption]. At this point, however, I see no reason to resist concluding that this method just *is* that which subserves mind reading. To deny this is to claim that there are two interpretive methods for attributing mental states, one subserving mind reading and one subserving confabulation. While this is possible, I find the alternative much more plausible.

Recall the importance of undermining [Auxiliary Assumption]. Without it, Engelbert and Carruthers cannot justify the predictive step of their argument. Specifically, because [Auxiliary Assumption] is false (and because the inner sense theory is compatible with its falsity), the inner sense theory does not predict that there should exist individuals who are incapable of self-attributing certain types of mental states (but not others). For this reason, contra Engelbert and Carruthers, the apparent absence of such individuals does not disconfirm the inner sense theory.

However, as I noted in section 3.3, there are other kinds of first-personal deficits/impairments, and thus other ways of instantiating the first, predic-

https://doi.org/10.1086/673967 Published online by Cambridge University Press

tive step of the two-step argument. In the remainder of this section, I will adduce various considerations that undermine several seemingly plausible ways of instantiating this predictive step.

6.1. Mind Reading Is Reliable. I argued a few paragraphs back that the method argued for in section 5 is likely that which subserves mind reading. I assume that mind reading others is a fairly reliable process; the deliverances of the mind-reading faculty are true often enough. Mind reading oneself should thus be at least as reliable. Indeed, there is reason to think that it should be even more reliable. As many have previously noted, there is a much greater bank of behavioral evidence when one's mind-reading target is oneself.[8]

Of course mind reading oneself delivers the wrong answers in confabulation studies, and this might appear to cast doubt upon the reliability of mind reading. But recall the reasonableness of the confabulated intentions in the Wegner and Wheatley study. They were perfectly reasonable, given the behavior and context of the participants. Presumably, the self-attributions made in confabulation studies are false because of the nature of the experimental designs. These studies create a mismatch between the true explanation of one's behavior and the mentalistic explanation of one's behavior that is seemingly most reasonable. In normal circumstances, this mismatch is presumably absent, and thus such self-attributions would most likely be true.

Accordingly, if the inner sense theory is true, those individuals with damaged inner sense will nevertheless have a reliable means of self-attributing mental states.[9] The inner sense theory thus does not predict individuals who unreliably self-attribute mental states.

6.2. Less Reliable Self-Attributions Would Likely Go Unnoticed. Consistent with the reliability of mind reading is the possibility that this method is less reliable than any alleged mechanism of inner sense. And, indeed, this is likely the case. After all, the inner sense theory likens self-knowledge to perceptual knowledge, which is quite reliable. Those with damaged inner sense should thus be expected to self-attribute mental states less reliably than those with unimpaired inner sense. Nevertheless, there are good reasons to think that this would go unnoticed.

8. Gilbert Ryle (1949/2000) famously made this point, although not in terms of 'mind reading'. More recently, Carruthers (2009, 2011) has used an updated version of this thought in defense of his account of self-knowledge.

9. If mind reading typically involves inner sense, then the reliability of mind reading in normal people cannot be transferred to cases where inner sense is impaired. However, as I argued in sec. 5.1, versions of the simulation theory that involve inner sense must nevertheless include nonsimulationist components that are capable on their own of generating reasonable mental state attributions. A decrease in reliability will be discussed shortly.

https://doi.org/10.1086/673967 Published online by Cambridge University Press

First, there is a presumed authority that we grant others with respect to their self-attributions. When what is at issue is the state of another's mind, we typically defer to the judgments of that individual. Consider, for example, the oddness of questioning another's claim to be depressed or to be thinking about one's day. Thus, many false self-attributions would likely go unnoticed.[10] Second, self-attributions can often be self-fulfilling. That is, even if one falsely self-attributes, for example, an intention to Φ, one might eventually come to form the intention to Φ in response to this self-attribution. In such cases, the (initially false) self-attribution will appear to be true.

Together, these points suggest that on the assumption that damage to one's inner sense should be expected to decrease the reliability of one's self-attributions, this decrease would likely go unnoticed. Accordingly, the inner sense theory does not predict that those with damaged inner sense should stand out in virtue of their less reliable self-attributions.

6.3. Mind Reading Oneself Need Not Feel Distinctively Third-Personal. I have argued that the method subserving mind reading could fill in for a damaged inner sense. Of course this method differs greatly from inner sense. While the latter is claimed to be direct and perception-like, the former is indirect and highly interpretive. Intuitively, then, the operation of these methods should feel quite differently to their subjects.

Research on so-called split-brain individuals is relevant here. In cases of severe epilepsy, some patients opt for cerebral commissurotomy, an operation that removes one's corpus callosum and anterior commissure. As a result, information received by the right hemisphere is closed off from the left hemisphere, and vice versa. Because the left hemisphere is verbal and the right hemisphere is nonverbal, speech for these patients is the sole product of the left hemisphere.

Michael Gazzaniga has conducted a number of fascinating studies on split-brain patients; see Gazzaniga (1995). In such studies, the patient is prompted to behave in a certain way due to information presented to his or her right hemisphere. For example, a patient's right hemisphere might receive the command to "get up and leave the room," causing the compliant patient to do just that. Due to the commissurotomy, the patient's left hemisphere is unaware of the true cause of the behavior. Interestingly, though, when asked to explain the behavior, the (left hemisphere of the) patient offers a mental explanation that is clearly confabulated. For example, the above

10. Deference is not an absolute rule, for we can imagine situations where such questioning is appropriate. The idea is rather that individuals are taken to be default authorities on their minds. Some philosophers, interpreting Wittgenstein (1953/2003), have suggested that such authority is partially constitutive of mental state self-attributions; see Wright (1989) for discussion of this view.

https://doi.org/10.1086/673967 Published online by Cambridge University Press

compliant patient might say "I was thirsty and wanted a drink." The explanation, although reasonable in light of the evidence available to (the left hemisphere of) the patient, is the product of the patient's interpretive mind-reading method.

So far, though, this is simply more evidence of confabulation. To see the significance of this research, consider the following from Gazzaniga. "While the patients possess at least some understanding of their surgery, they never say things like, 'Well I chose this because I have a split brain and the information went to the right, nonverbal hemisphere' . . . [they] view their responses as behaviors emanating from their own volitional selves, and as a result they incorporate these behaviors into a theory to explain why they behave as they do" (1995, 1394).

Apparently, these individuals do not detect their own confabulations. This is so despite their being aware that, due to their commissurotomies, they are likely to confabulate in certain situations. Indeed, as Carruthers (2009) reports (via communication with Gazzaniga), these individuals were often reminded of their situations during the experiments. He writes that "on a number of occasions testing was paused and the experimenter said something like, 'Joe, as you know, you have had this operation that sometimes will make it difficult for you to say what we show you over here left of fixation. You may find that your left hand [controlled by his right hemisphere] points to things for that reason, ok?' Joe assents, but then on the very next series he is back to the interpreter effect once again" (126). This strongly suggests that the method producing these confabulated explanations, although indirect and interpretive, does not feel that way to the patients.

I see no reason to resist extrapolating these results to normal, non-split-brain individuals. Denying this is to suppose, implausibly, that a commissurotomy changes the feeling of the indirect, interpretive method housed in one's left hemisphere. Individuals forced to self-attribute mental states via their mind-reading method should thus not be assumed to notice this fact. Like Gazzaniga's split-brain patients, many of their self-attributions may nevertheless feel direct and noninterpretive. Perhaps this should not be too surprising. After all, normal participants in confabulation studies presumably do not (always) feel as though they are interpreting themselves.[11] I conclude, then, that the inner sense theory does not predict that those with damaged inner sense should feel as though they have only an indirect, third-personal access to their minds.

11. Indeed, mind reading others can often feel direct and noninterpretive. Sometimes one can simply "see" that another is sad, or desiring food, etc.

https://doi.org/10.1086/673967 Published online by Cambridge University Press

7. Implications for the Second Step. The previous section contains three arguments, each of which shows that an apparent prediction of the inner sense theory is not a prediction of that theory after all. In addition, I earlier blocked the prediction from Engelbert and Carruthers's argument. These arguments accomplish more than the mere blocking of various ways of instantiating the predictive step of the two-step argument. Collectively, they show that individuals with damaged inner sense should appear to be quite normal to both themselves and others, if they have intact mind reading. First, they should be able to self-attribute any kind of mental state that they can attribute to others. Second, they should be able to do so reliably. Third, although they should do so less reliably than their unimpaired peers, this difference should go unnoticed. Finally, there is evidence that their self-attributions should feel direct and first-personal.

These points significantly complicate the second step of the two-step argument, namely, the step claiming that the prediction is or is not borne out. Because individuals with damaged inner sense will appear to be quite normal, those first-personal deficits/impairments that are legitimate predictions of the inner sense theory will be quite subtle and consequently difficult to detect. Justifying the second step of the two-step argument will thus be very difficult. Take, for example, the legitimate prediction (noted in sec. 6.2) that those with damaged inner sense should self-attribute mental states less reliably than their unimpaired peers. Imagine the difficulties with trying to establish the existence of such individuals. I suppose there is no *in principle* difficulty here, but, for the reasons given in section 6.2, the practical difficulties would be significant.

7.1. Nichols and Stich's Argument. Incidentally, this discussion is relevant to Nichols and Stich's argument. Unfortunately, I have the space here to make only the briefest remark. They claim that the prediction from their argument is borne out by schizophrenics with 'passivity experiences', that is, delusions of control, thought insertion, and thought withdrawal. Such symptoms are anything but subtle, however! If what I have argued so far is correct, then one should be suspicious of Nichols and Stich's claim that these symptoms are due to damaged inner sense (coincident with intact mind reading). And, indeed, others have persuasively criticized their claim; see, for example, Wiffen and David (2009) and Carruthers (2011).

8. Conclusion. The arguments from the previous four sections show that successfully testing the inner sense theory via the kind of two-step argument discussed in this article is quite difficult, requiring much more care and thoroughness than perhaps was previously thought. The existence of the method that subserves confabulation (and mind reading) seriously complicates matters, presenting difficulties at each of the two steps.

https://doi.org/10.1086/673967 Published online by Cambridge University Press

REFERENCES

Armstrong, David M. 1981. *The Nature of Mind and Other Essays*. Ithaca, NY: Cornell University Press.

Brasil-Neto, Joaquim, Alvaro Pascual-Leone, Josep Valls-Sole, Leonardo Cohen, and Mark Halett. 1992. "Focal Transcranial Magnetic Stimulation and Response Bias in a Forced Choice Task." *Journal of Neurology, Neurosurgery, and Psychiatry* 55:964–66.

Carruthers, Peter. 2009. "How We Know Our Own Minds: The Relationship between Mindreading and Metacognition." *Behavioral and Brain Sciences* 32:121–82.

———. 2011. *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Eagly, Alice H., and Shelly Chaiken. 1993. *The Psychology of Attitudes*. Fort Worth, TX: Harcourt Brace Jovanovich.

Engelbert, Mark, and Peter Carruthers. 2010. "Introspection." *WIREs Cognitive Science* 1:245–53.

Gazzaniga, Michael. 1995. "Consciousness and the Cerebral Hemisphere." In *The Cognitive Neurosciences*, ed. Michael Gazzaniga, 1391–1400. Cambridge, MA: MIT Press.

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

Gordon, Robert. 1995. "Simulation without Introspection or Inference from Me to You." In *Mental Simulation*, ed. Martin Davies and Tony Stone, 53–67. Cambridge, MA: Blackwell.

Lycan, William G. 1996. *Consciousness and Experience*. Cambridge, MA: Bradford.

Nichols, Shaun, and Stephen Stich. 2003. *Mindreading: An Integrated Account of Pretense, Self-Awareness, and Understanding of Other Minds*. Oxford: Oxford University Press.

Nisbett, Richard, and Timothy Wilson. 1977. "Telling More than We Can Know." *Psychological Review* 84:231–95.

Ryle, Gilbert. 1949/2000. *The Concept of Mind*. Repr. Chicago: University of Chicago Press.

Wegner, David, and Thalia Wheatley. 1999. "Apparent Mental Causation: Sources of the Experience of the Will." *American Psychologist* 54:480–91.

Wiffen, Ben, and Anthony David. 2009. "Metacognition, Mindreading, and Insight in Schizophrenia." *Behavioral and Brain Sciences* 32:161–62.

Wittgenstein, Ludwig. 1953/2003. *The Philosophical Investigations*. Trans. G. E. M Anscombe. Repr. Oxford: Blackwell.

Wright, Crispin. 1989. "Wittgenstein's Later Philosophy of Mind: Sensation, Privacy, and Intention." *Journal of Philosophy* 86:622–34.

https://doi.org/10.1086/673967 Published online by Cambridge University Press