Stanford Encyclopedia of Philosophy

Introspection

First published Tue Feb 2, 2010; substantive revision Thu Apr 24, 2014

Introspection, as the term is used in contemporary philosophy of mind, is a means of learning about one's own currently ongoing, or perhaps very recently past, mental states or processes. You can, of course, learn about your own mind in the same way you learn about others' minds—by reading psychology texts, by observing facial expressions (in a mirror), by examining readouts of brain activity, by noting patterns of past behavior—but it's generally thought that you can also learn about your mind *introspectively*, in a way that no one else can. But what exactly is introspection? No simple characterization is widely accepted.

Introspection is a key concept in epistemology, since introspective knowledge is often thought to be particularly secure, maybe even immune to skeptical doubt. Introspective knowledge is also often held to be more immediate or direct than sensory knowledge. Both of these putative features of introspection have been cited in support of the idea that introspective knowledge can serve as a ground or foundation for other sorts of knowledge.

Introspection is also central to philosophy of mind, both as a process worth study in its own right and as a court of appeal for other claims about the mind. Philosophers of mind offer a variety of theories of the nature of introspection; and philosophical claims about consciousness, emotion, free will, personal identity, thought, belief, imagery, perception, and other mental phenomena are often thought to have introspective consequences or to be susceptible to introspective verification. For similar reasons, empirical psychologists too have discussed the accuracy of introspective judgments and the role of introspection in the science of the mind.

1. General Features of Introspection

1.1 Necessary Features of an Introspective Process

Introspection is generally regarded as a process by means of which we learn about our own currently ongoing, or very recently past, mental states or processes. Not all such processes are introspective, however: Few would say that you have introspected if you learn that you're angry by seeing your facial expression in the mirror. However, it's unclear and contentious exactly what more is required for a process to qualify as introspective. A relatively restrictive account of introspection might require introspection to involve attention to and direct detection of one's ongoing mental states; but many philosophers think attention to or direct detection of mental states is impossible or at least not present in many paradigmatic instances of introspection.

For a process to qualify as “introspective” as the term is ordinarily used in contemporary philosophy of mind, it must minimally meet the following three conditions:

1. *The mentality condition*: Introspection is a process that generates, or is aimed at generating, knowledge, judgments, or beliefs about *mental* events, states, or processes, and not about affairs outside one's mind, at least not directly. In this respect, it is different from sensory processes that normally deliver information about outward events or about the non-mental aspects of the individual's body. The border between introspective and non-introspective knowledge can begin

to seem blurry with respect to bodily self-knowledge such as proprioceptive knowledge about the position of one's limbs or nociceptive knowledge about one's pains. But it seems that in principle the introspective part of such processes, pertaining to judgments about one's mind—e.g., that one has the feeling as though one's arms were crossed or of toe-ishly located pain—can be distinguished from the non-introspective judgment that one's arms are in fact crossed or one's toe is being pinched.

2. *The first-person condition*: Introspection is a process that generates, or is aimed at generating, knowledge, judgments, or beliefs about *one's own mind only* and no one else's, at least not directly. Any process that in a similar manner generates knowledge of one's own and others' minds is by that token not an introspective process. (Some philosophers have contemplated peculiar or science fiction cases in which we might introspect the contents of others' minds directly—for example in telepathy or when two individuals' brains are directly wired together—but the proper interpretation of such cases is disputable see, e.g., Gertler 2000.)

3. *The temporal proximity condition*: Introspection is a process that generates knowledge, beliefs, or judgments about one's *currently ongoing* mental life only; or, alternatively (or perhaps in addition) *immediately past* (or even future) mental life, within a certain narrow temporal window (sometimes called the specious present; see the entry on the experience and perception of time). You may know that you were thinking about Montaigne yesterday during your morning walk, but you cannot know that fact by current introspection alone—though perhaps you can know introspectively that you currently have a vivid memory of having thought about Montaigne. Likewise, you cannot know by introspection alone that you will feel depressed if your favored candidate loses the election in November—though perhaps you can know introspectively what your current attitude is toward the election or what emotion starts to rise in you when you consider the possible outcomes. Whether the target of introspection is best thought of as one's current mental life or one's immediately past mental life may depend on one's model of introspection: On self-detection models of introspection, according to which introspection is a causal process involving the detection of a mental state (see Section 2.2 below), it's natural to suppose that a brief lapse of time will transpire between the occurrence of the mental state that is the introspective target and the final introspective judgment about that state, which invites (but does not strictly imply) the idea that introspective judgments generally pertain to immediately past states. On self-shaping and self-fulfillment models of introspection, according to which introspective judgments create or embed the very state introspected (see Sections 2.3.1 and 2.3.2 below), it seems more natural to think that the target of introspection is one's current mental life or perhaps even the immediate future.

Few contemporary philosophers of mind would call a process "introspective" if it does not meet some version of the three conditions above, though in ordinary language the temporal proximity condition may sometimes be violated. (For example, in ordinary speech we might describe as "introspective" a process of thinking about why you abandoned a relationship last month or whether you're really as kind to your children as you think you are.) However, many philosophers of mind will resist calling a process that meets these three conditions "introspective" unless it also meets some or all of the following three conditions:

4. *The directness condition*: Introspection yields judgments or knowledge about one's own current

mental processes relatively *directly* or *immediately*. It's difficult to articulate exactly what directness or immediacy involves in the present context, but some examples should make the import of this condition relatively clear. Gathering sensory information about the world and then drawing theoretical conclusions based on that information should not, according to this condition, count as introspective, even if the process meets the three conditions above. Seeing that a car is twenty feet in front of you and then inferring from that fact about the external world that you are having a visual experience of a certain sort does not, by this condition, count as introspective. However, as we will see in Section 2.3.4 below, those who embrace transparency theories of introspection may reject at least strong formulations of this condition.

5. *The detection condition*: Introspection involves some sort of *attunement to* or *detection of* a *pre-existing* mental state or event, where the introspective judgment or knowledge is (when all goes well) *causally* but not *ontologically* dependent on the target mental state. For example, a process that involved creating the state of mind that one attributes to oneself would not be introspective, according to this condition. Suppose I say to myself in silent inner speech, “I am saying to myself in silent inner speech, ‘haecceities of applesauce’”, without any idea ahead of time how I plan to complete the embedded quotation. Now, what I say may be true, and I may know it to be true, and I may know its truth (in some sense) directly, by a means by which I could not know the truth of anyone else's mind. That is, it may meet all the four conditions above and yet we may resist calling such a self-attribution introspective. Self-shaping (Section 2.3.2 below), expressivist (Section 2.3.3 below), and transparency (Section 2.3.4 below) accounts of self-knowledge emphasize the extent to which our self-knowledge often does not involve the detection of pre-existing mental states; and because something like the detection condition is implicitly or explicitly accepted by many philosophers, some philosophers (including some but not all of those who endorse self-shaping, expressivist, and/or transparency views) would regard it as inappropriate to regard such accounts of self-knowledge as accounts of *introspection* proper.

6. *The effort condition*: Introspection is not *constant, effortless, and automatic*. We are not every minute of the day introspecting. Introspection involves some sort of special reflection on one's own mental life that differs from the ordinary un-self-reflective flow of thought and action. The mind may monitor itself regularly and constantly without requiring any special act of reflection by the thinker—for example, at a non-conscious level certain parts of the brain or certain functional systems may monitor the goings-on of other parts of the brain and other functional systems, and this monitoring may meet all five conditions above—but this sort of thing is not what philosophers generally have in mind when they talk of introspection. However, this condition, like the directness and detection conditions, is not universally accepted. For example, philosophers who think that conscious experience requires some sort of introspective monitoring of the mind and who think of conscious experience as a more or less constant feature of our lives may reject the effort condition (Armstrong 1968, 1999; Lycan 1996).

Though not all philosophical accounts that are put forward by their authors as accounts of “introspection” meet all of conditions 4–6, most meet at least two of those. Because of differences in the importance accorded to conditions 4–6, it is not unusual for authors with otherwise similar accounts of self-knowledge to differ in their willingness to describe their accounts as accounts of “introspection”.

1.2 The Targets of Introspection

Accounts of introspection differ in what they treat as the proper *targets* of the introspective process. No major contemporary philosopher believes that all of mentality is available to be discovered by introspection. For example, the cognitive processes involved in early visual processing and in the detection of phonemes are generally held to be introspectively impenetrable and nonetheless (in some important sense) mental (Marr 1983; Fodor 1983). Many philosophers also accept the existence of unconscious beliefs or desires, in roughly the Freudian sense, that are not introspectively available (e.g., Gardner 1993; Velleman 2000; Moran 2001; Wollheim 2003; though see Lear 1998). Although in ordinary English usage we sometimes say we are “introspecting” when we reflect on our character traits, contemporary philosophers of mind generally do not believe that we can directly introspect character traits in the same sense in which we can introspect some of our other mental states (especially in light of research suggesting that we sometimes have poor knowledge of our traits, reviewed in Taylor and Brown 1988; Paulhus and John 1998; Vazire 2010).

The two most commonly cited classes of introspectible mental states are *attitudes*, such as beliefs, desires, evaluations, and intentions, and *conscious experiences*, such as emotions, images, and sensory experiences. (These two groups may not be wholly, or even partially, disjoint: Depending on other aspects of her view, a philosopher may regard some or all conscious experiences as involving attitudes, and/or she may regard attitudes as things that are or can be consciously experienced.) It of course does not follow from the fact (if it is a fact) that some attitudes are introspectible that all attitudes are, or from the fact that some conscious experiences are introspectible that all conscious experiences are. Some accounts of introspection focus on attitudes (e.g., Nichols and Stich 2003), while others focus on conscious experiences (e.g., Hill 1991; Goldman 2006; Schwitzgebel 2012); and it is sometimes unclear to what extent philosophers intend their remarks about the introspection of one type of target to apply to the other type. There is no guarantee that the same mechanism or process is involved in introspecting all the different potential targets.

Generically, this article will describe the targets of introspection as *mental states*, though in some cases it may be more apt to think of the targets as processes rather than states. Also, in speaking of the targets of introspection as *targets*, no presupposition is intended of a self-detection view of introspection as opposed to a self-shaping or containment or expressivist view (see Section 2 below). The targets are simply the states self-ascribed as a consequence of the introspective process if the process works correctly, or if the introspective process fails, the states that would have been self-ascribed.

1.3 The Products of Introspection

Though philosophers have not explored the issue very thoroughly, accounts also differ regarding the *products* of introspection. Most philosophers hold that introspection yields something like beliefs or judgments about one's own mind, but others prefer to characterize the products of introspection as “thoughts”, “representations”, “awareness”, or the like. For ease of exposition, this article will describe the products of the introspective process as judgments, without meaning to beg the question against competing views.

2. Introspective Versus Non-Introspective Accounts of Self-Knowledge

This section will outline several approaches to self-knowledge. Not all deserve to be called introspective, but an understanding of introspection requires an appreciation of this diversity of approaches—some for the sake of the contrast they provide to introspection proper and some because it's disputable whether they should be classified as introspective. These approaches are not exclusive. Surely there is more than one process by means of which we can obtain self-knowledge. Unavoidably, some of the same territory covered here is also covered, rather differently, in the entry on self-knowledge.

2.1 Self/Other Parity Accounts

Symmetrical or *self/other parity* accounts of self-knowledge treat the processes by which we acquire knowledge of our own minds as essentially the same as the processes by which we acquire knowledge of other people's minds. A simplistic version of this view is that we know both our own minds and the minds of others only by observing outward behavior. On such a view, introspection strictly speaking is impossible, since the first-person condition on introspection (condition 2 in Section 1.1) cannot be met: There is no distinctive process that generates knowledge of one's own mind only. Twentieth-century behaviorist principles tended to encourage this view, but no prominent treatment of self-knowledge accepts this view in its most extreme and simple form. Advocates of parity accounts sometimes characterize our knowledge of our own minds as arising from "theories" that we apply equally to ourselves and others (as in Nisbett and Ross 1980; Gopnik 1993a, 1993b). Consequently, this approach to self-knowledge is sometimes called the *theory theory*.

2.1.1 Behavioral Observation Accounts

Among leading researchers, Bem (1972) perhaps comes closest to a simple self/other parity view, arguing on the basis of psychological research that our knowledge of the "internal states" of both self and other derives largely from the same types of behavioral evidence and employs the same principles of inference. We notice how we behave, and then we infer the attitudes evidenced by those behaviors—and we do so even when we actually lack the ascribed attitude. For example, Bem cites classic research in social psychology suggesting that when induced to perform an action for a small reward, people will attribute to themselves a more positive attitude toward that action than when they are induced by a large reward (Festinger and Carlsmith 1959; see also Section 4.2.2 below). When we notice ourselves doing something with minimal compensation, we infer a positive attitude toward that activity, just as we would if we saw someone else perform the same activity with minimal compensation. Likewise, we might know we like Thai food because we've noticed that we sometimes drive all the way across town to get it; we might know that we're happy because we see or feel ourselves smiling. Bem argues that social psychology has consistently failed to show that we have any appreciable access to private information that might tell against such externally-driven self-attributions. On Bem's view, if we are better at discerning our own motives and attitudes, it's primarily because we have observed more of our own behavior than of anyone else's.

2.1.2 Theory Theory Accounts

Nisbett, Wilson, and their co-authors (Nisbett and Bellows 1977; Nisbett and Wilson 1977; Nisbett and Ross 1980; Wilson 2002) similarly argue for self/other parity in our knowledge of the bases or causes of our own and others' attitudes and behavior, describing cases in which people seem to show poor knowledge of these bases or causes. For example, people queried in a suburban shopping center about why they chose a particular pair of stockings appeared to be ignorant of the influence of position on that choice, including explicitly denying that influence when it was suggested to them. People asked to rate various traits of supposed job applicants were unaware that their judgments of the applicant's flexibility were greatly influenced by having been told that the applicant had spilled coffee during the job interview (see also Section 4.2.2 below). In such cases, Nisbett and his co-investigators found that subjects' descriptions of the causal influences on their own behavior closely mirrored the influences hypothesized by outside observers. From this finding, they infer that the same mechanism drives the first-person and third-person attributions, a mechanism that that does not involve any special private access to the real causes of one's attitudes and behavior and instead relies heavily on intuitive psychological theories.

Gopnik (1993a, 1993b; Gopnik and Meltzoff 1994) deploys developmental psychological evidence to support a parity theory of self-knowledge. She points to evidence that for a wide variety of mental states, including believing, desiring, and pretending, children develop the capacity to ascribe those states to themselves at the same age they develop the capacity to ascribe those states to others. For example, children do not seem to be able to ascribe to themselves past false beliefs (after having been tricked by the experimenter) any earlier than they can ascribe false beliefs to other people. This appears to be so even when that false belief is in the very recent past, having only just been revealed to be false. According to Gopnik, this pervasive parallelism shows that we are not given direct introspective access to our beliefs, desires, pretenses, and the like. Rather, we must develop a “theory of mind” in light of which we interpret evidence underwriting our self-attributions. The appearance of the immediate givenness of one's mental states is, Gopnik suggests, merely an “illusion of expertise”: Experts engage in all sorts of tacit theorizing that they don't recognize as such—the expert chess player for whom the strength of a move seems simply visually given, the doctor who immediately intuits cancer in a patient. Since we are all experts at mental state attribution, we don't recognize the layers of theory underwriting the process.

2.1.3 Restrictions on Parity

The empirical evidence behind self/other parity views remains contentious (White 1988; Nichols and Stich 2003). Furthermore, though Bem, Nisbett, Wilson, and Gopnik all stress the parallelism between mental state attribution to oneself and others and the inferential and theoretical nature of such attributions, they all also leave some room for a kind of self-awareness different in kind from the awareness one has of others' mental lives. Thus, none endorses a *purely* symmetrical or self/other parity view. Bem acknowledges that the parallelism only holds “to the extent that internal cues are weak, ambiguous, or uninterpretable” (1972, 5). With this caveat in mind, he states that our self-knowledge is “partially” based on external cues. Nisbett and Wilson stress that we lack access only to the “processes” or causes *underlying* our behavior and attitudes. Our attitudes themselves and our current sensations, they say, can be known with “near certainty” (1977, 255; though contrast

Nisbett and Ross 1980, 200–202, which seems sympathetic to Bem's skepticism about special access even to our attitudes). Gopnik allows that we "may be well equipped to detect certain kinds of internal cognitive activity in a vague and unspecified way", and that we have "genuinely direct and special access to certain kinds of first-person evidence [which] might account for the fact that we can draw some conclusions about our own psychological states when we are perfectly still and silent", though we can "override that evidence with great ease" (1993a, 11–12). Ryle (1949) similarly stresses the importance of outward behavior in the self-attribution of mental states while acknowledging the presence of "twinges", "thrills", "tickles", and even "silent soliloquies", which we know of in our own case and that do not appear to be detectable by observing outward behavior. However, none of these authors develops an account of this apparently more direct self-knowledge. Their theories are consequently incomplete. Regardless of the importance of behavioral evidence and general theories in driving our self-attributions, in light of the considerations that drive Bem, Nisbett, Wilson, Gopnik, and Ryle to these caveats, it is probably impossible to sustain a view on which there is complete parity between first- and third-person mental state attributions. There must be some sort of introspective, or at least uniquely first-person, process.

Self/other parity views can also be restricted to particular subclasses of mental states: Any mental state that can only be known by cognitive processes identical to the processes by which we know about the same sorts of states in other people is a state to which we have no distinctively introspective access. States for which parity is often asserted include personality traits, unconscious motives, early perceptual processes, and the bases of our decisions (see Section 4.2.1 below for more on this). We learn about these states in ourselves, perhaps, in much the same way we learn about such states in other people. Carruthers (2011; see also Section 4.2.2 below) presents a case for parity of access to propositional attitudes like belief and desire (in contrast to inner speech, visual imagery, and the like, which he holds to be introspectible).

2.2 Self-Detection Accounts

Etymologically, the term "introspection"—from the Latin "looking into"—suggests a perceptual or quasi-perceptual process. Locke writes that we have a faculty of "Perception of the Operation of our own Mind" which, "though it be not Sense, as having nothing to do with external Objects; yet it is very like it, and might properly enough be call'd internal Sense" (1690/1975, 105, italics suppressed). Kant (1781/1997) says we have an "inner sense" by which we learn about mental aspects of ourselves that is in important ways parallel to the "outer sense" by which we learn about outer objects.

But what does it mean to say that introspection is like perception? In what respects? As Shoemaker (1994a, 1994b, 1994c) points out, in a number of respects introspection is plausibly *unlike* perception. For example, introspection does not involve a dedicated organ like the eye or ear (though as Armstrong 1968 notes, neither does bodily proprioception). Both friends and foes of self-detection accounts have tended to agree that introspection does not involve a distinctive phenomenology of "introspective appearances" (Shoemaker 1994a, 1994b, 1994c; Lycan 1996; Rosenthal 2001; Siewert 2012): The visual experience of redness has a distinctive sensory quality or phenomenology that would be difficult or impossible to convey to a blind person; analogously for the olfactory experience of smelling a banana, the auditory experience of hearing a pipe organ, the experience of touching something painfully hot. To be analogous to sensory experience in this respect, introspection would have to generate an analogously distinctive phenomenology—some quasi-sensory phenomenology in addition to, say, the visual phenomenology of seeing red that is the phenomenology of the *introspective appearance* of the visual phenomenology of seeing red. This would seem to require two layers of appearance in introspectively attended sensory perception: a visual appearance of the outward object and an introspective appearance of that visual appearance. (This isn't to say, however, that introspection, or at least conscious introspection, doesn't involve some sort of “cognitive phenomenology”—if there is such a thing—of the sort that accompanies conscious thoughts in general: See Bayne and Montague, eds., 2011.)

Contemporary proponents of quasi-perceptual models of introspection concede the existence of such disanalogies (e.g., Lycan 1996). We might consider an account of introspection to be quasi-perceptual, or less contentiously to be a “self-detection” account, if it meets the first five conditions described in Section 1.1—that is, the mentality condition, the first-person condition, the temporal proximity condition, the directness condition, and the detection condition. One aspect of the detection condition deserves special emphasis here: that detection requires the ontological independence of the target mental state and the introspective judgment—the two states will be causally connected (assuming that all has gone well) but not *constitutively* connected. (Shoemaker (1994a, 1994b, 1994c) calls models of self-knowledge that meet this aspect of the detection condition “broad perceptual” models.) Maybe on a liberal understanding of “detection” that does not require ontological independence, containment or other accounts of introspection (see Section 2.3.1 below) might qualify as involving “detection”. However, that is not how “detection” is being used in the present taxonomy.

Self-detection accounts of self-knowledge seem to put introspection epistemically on a par with sense perception. To many philosophers, this has seemed a deficiency in these accounts. A long and widespread philosophical tradition holds that self-knowledge is epistemically special, that we have specially “privileged access” to—perhaps even infallible or indubitable knowledge of—at least some portion of our mentality, in a way that is importantly different in kind from our knowledge of the world outside us (see Section 4 below). Both self/other parity accounts (Section 2.1 above) and self-detection accounts (this section) of self-knowledge either deny any special epistemic privilege or characterize that privilege as similar to the privilege of being the only person to have an extended view of an object or a certain sort of sensory access to that object. Other accounts of self-knowledge to be discussed later in Section 2.3 are more readily compatible with, and often to some extent driven by, more robust notions of the epistemic differences between self-knowledge and knowledge of environmental objects.

2.2.1 Simple Monitoring Accounts

Armstrong (1968, 1981, 1999) is perhaps the leading defender of a quasi-perceptual, self-detection account of introspection. He describes introspection as a “self-scanning process in the brain” (1968, 324), and he stresses what he sees as the important ontological distinction between the state of awareness produced by the self-scanning procedure and the target mental state of which one is aware by means of that scanning—the distinction, for example, between one's pain and one's introspective awareness of that pain.

Armstrong also appears to hold that the quasi-perceptual introspective process proceeds at a fairly low level cognitively—quick and simple, typically without much interference by or influence from other cognitive or sensory processes. He describes introspection as “completely non-inferential”, similar to the simple detection of pressure on one's back (1968, 97), and he says it can be (and presumably typically is) continuous and “reflex”, involving no more than keeping “a watching brief on our own current mental contents, but without making much of a deal of it” (1999, 115). Since Armstrong allows that inferences are often non-conscious, based on sensory or other cues that the inferring person cannot herself discern, his claim that the introspective process is non-inferential is a substantial commitment to the simplicity of the process. He contrasts this reflexive self-monitoring with more sophisticated acts of deliberate introspection which he thinks are also possible (1999, 114). Note that in calling reflexive self-monitoring “introspection”, Armstrong violates the effort condition from Section 1.1, which requires that introspection not be constant and automatic. Lycan (1996) endorses a similar view, though unlike Armstrong, Lycan characterizes introspection as involving *attentional* mechanisms, thus presumably treating introspection as more demanding of cognitive resources (though still perhaps nearly constant).

Nichols and Stich (2003) employ a model of the mind on which having a propositional attitude such as a belief or desire is a matter of having a representation stored in a functionally-defined (and metaphorical) “belief box” or “desire box” (see also the entries on belief and functionalism). On their account, self-awareness of these attitudes typically involves the operation of a simple “Monitoring Mechanism” that merely takes the representations from these boxes, appends an “I believe that …”, “I desire that …”, or whatever (as appropriate) to that representation, and adds it back into the belief box. For example, if I desire that my father flies to Hong Kong on Sunday, the Monitoring Mechanism can copy the representation in my desire box with the content “my father flies to Hong Kong on Sunday” and produce a new representation in my belief box—that is, create a new belief—with the content “I desire that my father flies to Hong Kong on Sunday”. Nichols and Stich also propose an analogous but somewhat more complicated mechanism (they leave the details unspecified) that takes percepts as its input and produces beliefs about those percepts as its output.

Nichols and Stich emphasize that this Monitoring Mechanism does not operate in isolation, but often co-operates or competes with a second means of acquiring self-knowledge, which involves deploying theories along the lines suggested by Gopnik (see Section 2.1.2 above). They offer a “double dissociation” argument for this view. That is, they present, on the one hand, cases which they interpret as cases showing a breakdown in the Monitoring Mechanism, while the capacity for theoretical inference about the mind remains intact and, on the other hand, cases in which the capacity for theoretical inference about the mind is impaired but the Monitoring Mechanism continues to function normally, suggesting that theoretical inference and self-monitoring are distinct and separable processes. Nichols and Stich argue that autistic people have very poor theoretical knowledge of the mind, as suggested by their very poor performance in “theory of mind” tasks (tasks like assessing when someone will have a false belief), and yet they succeed in monitoring their mental states as shown by their ability to describe their mental states in autobiographies and other forms of self-report. Conversely, Nichols and Stich argue that schizophrenic people remain excellent theorizers about mental states but monitor their own mental states very poorly—for example, when they fail to recognize certain actions as their own and struggle to report, or deny the existence of, ongoing thoughts.

2.2.2 Multi-Process Monitoring Accounts

Goldman (2006) criticizes the account of Nichols and Stich (see Section 2.2.1 above) for not describing how the Monitoring Mechanism detects the attitude type of the representation (belief, desire, etc.). If talk of “belief boxes” and the like is shorthand for talk of functional role (as Nichols and Stich say), then the Monitoring Mechanism must somehow detect the functional role of the detected representation. But functional role is a matter of what is apt to cause a particular mental state and what that mental state is apt to cause (see the entry on functionalism), and Goldman argues that a simple mechanism could not discern such dispositional and relational facts (though Nichols and Stich might be able to avoid this concern by describing introspection as involving not just one but rather a cluster of similar mechanisms: 2003, 162). Goldman also argues that the Nichols and Stich account leaves unclear how we can discern the strength or intensity of our beliefs, desires, and other propositional attitudes.

Goldman's positive account starts with the idea that introspection is a quasi-perceptual process that involves attention: “Attention seems to act like an orienting organ in introspection, analogous to the shift of eye gaze or the sniffing of the nose” (2006, 244). Individual attended mental states are then classified into broad categories (similarly, in visual perception we can classify seen objects into broad categories). However, on Goldman's view this process can only generate introspective knowledge of the general *types* of mental states (such as belief, happiness, bodily sensation) and some properties of those mental states (such as degree of confidence for belief, and “a multitude of finely delineated categories” for bodily sensation). Specific contents, especially of attitudes like belief, are too manifold, Goldman suggests, for pre-existing classificational categories to exist for each one. Rather, we represent the specific content of such mental states by “redeploying” the representational content of the mental state, that is, simply copying the content of the introspected mental state into the content of the introspective belief or judgment (somewhat like in the Nichols and Stich account). Finally, Goldman argues that some mental states require “translation” into the mental code appropriate to belief if they are to be introspected. Visual representations, he suggests, have a different format or mental code than beliefs, and therefore cognitive work will be necessary to translate the fine-grained detail of visual experience into mental contents that can be believed introspectively.

Hill (1991, 2009) also offers a multi-process self-detection account of introspection. Like Goldman, Hill sees attention (in some broad, non-sensory sense) as central to introspection, though he also allows for introspective awareness without attention (1991, 117–118). Hill emphasizes dissimilarities between introspection and perception, while retaining a broadly self-detection account. Hill (2009) argues that introspection is a process that produces *judgments about*, rather than perceptual awareness of, the target states, and suggests that the processes that generate these judgments vary considerably, depending on the target state, and are often complex. For example, judgments about enduring beliefs and desires must, he says, involve complex procedures for searching “vast and heterogeneous” long-term memory stores. Central to Hill's (1991) account is an emphasis on the capacity of introspective attention to transform—especially to amplify and enrich, even to create—the target experience. In this respect Hill argues that the introspective act differs from the paradigmatic observational act which does not transform the object perceived (though of course both scientific and ordinary—especially gustatory—observation can affect what is perceived); and thus Hill's account contains a “self-fulfillment” or “self-shaping” aspect in the sense of Section 2.3.1 and Section 2.3.2 below, and only qualifiedly and conditionally meets the detection condition on accounts of introspection as described in Section 1.1 above—the condition that introspection involves attunement to or detection of a pre-existing mental state or event.

Like Hill, Prinz (2004) argues that introspection must involve multiple mechanisms, depending both on the target states (e.g., attitudes vs. perceptual experiences) and the particular mode of access to those states. Access might involve controlled attention or it might be more of a passive noticing; it might involve the verbal "captioning" or labeling of experiences or it might involve the kind of non-verbal access that even monkeys have to their mental states. Prinz (2007) sharply distinguishes between the *conceptual classification* of our conscious experiences into various types that can be recognized and re-identified over time—classifications which he thinks must necessarily be somewhat crude—and non-conceptual knowledge of ongoing conscious experiences attained by "pointing" at them with attention. The latter type of knowledge, Prinz argues, is much more detailed and finely structured than the former but cannot be expressed or retained over time. Prinz also follows Hill in emphasizing that introspection often intensifies or otherwise modifies the target experience. In such cases, Prinz argues, introspective "access" is only access in an attenuated sense.

2.3 Introspection Without Self-Detection?

There are several ways to generate judgments, or at least statements, about one's own current mental life—self-ascriptions, let's call them—that are reliably true though they do not involve the detection of a pre-existing state. Consider the following four types of case:

A. *Automatically self-fulfilling self-ascriptions*: I think to myself, "I am thinking". Or: I judge that I am making a judgment about my own mental life. Or: I say to myself in inner speech "I am saying to myself in inner speech: 'blu-bob'". Such self-ascriptions are automatically self-fulfilling. Their existence conditions are a subset of their truth conditions.

B. *Self-ascriptions that prompt self-shaping*: I declare that I have a mental image of a pink elephant. At the same time I make this declaration, I deliberately cause myself to form the mental image of a pink elephant. Or: A man uninitiated in romantic love declares to a prospective lover that he is the kind of person who sends flowers to his lovers. At the same time he says this, he successfully resolves to be the kind of person who sends flowers to his lovers. The self-ascription either precipitates a change or buttresses what already exists in such a way as to make the self-ascription accurate. In these cases, unlike the cases described in (A), some change or self-maintenance is necessary to render the self-ascription true, beyond the self-ascriptional event itself.

C. *Accurate self-ascription through self-expression*: I learn to say "I'm in pain!" instead of "ow!" as an automatic, unreflective response to painful stimuli. Or: I use the self-attributive sentence "I believe Russell changed his mind about pacifism" simply as a cautious way of expressing the belief that Russell changed his mind about pacifism, this expression being the product of reflecting upon Russell rather than a product of reflection upon my own mind. Self-expressions of this sort are assumed here to flow naturally from the states expressed in roughly the same way that facial expressions and non-self-attributive verbal expressions flow naturally from those same states—that is, without being preceded by any attempt to detect the state self-ascribed.

D. *Self-ascriptions derived from judgments about the outside world*: From the non-self-attributive fact that Stanford is south of Berkeley I derive the self-attributive conclusion that I believe that Stanford is south of Berkeley. Or: From the non-self-attributive fact that it would be good to go

> to home now, I derive the self-attributive judgment that I want to go home now. These derivations may be inferences, but if so, such inferences require no specific premises about ongoing mental states. Perhaps one embraces a general inference principle like "from *P*, it is permissible to derive *I believe that P*", or "normally, if something is good, I want it".

The following accounts of self-knowledge all take advantage of one or more of these facts about self-ascription. Because these ways of obtaining self-knowledge all violate the detection condition on introspection (condition 5 in Section 1.1 above), and because philosophers are divided about whether methods of obtaining self-knowledge that violate that condition count as *introspective* methods strictly speaking, philosophers are divided about whether accounts of self-knowledge of the sort described in this section should be regarded as accounts of introspection.

2.3.1 Self-Fulfillment and Containment

An emphasis on infallible knowledge through self-fulfilling self-ascriptions goes back at least to Augustine (c. 420 C.E./1998) and is most famously deployed by Descartes in his *Discourse on Method* (1637/1985) and *Meditations* (1641/1984), where he takes the self-fulfilling thought that he is thinking as indubitably true, immune to even the most radical skepticism, and a secure ground on which to build further knowledge.

Contemporary self-fulfillment accounts tend to exploit the idea of *containment*. In a 1988 essay, Burge writes:

> When one knows one is thinking that *p*, one is not taking one's thought (or thinking) that *p* merely as an object. One is thinking that *p* in the very event of thinking knowledgeably that one is thinking it. It is thought and thought about in the same mental act. (654)

This is the case, Burge argues, because "by its reflexive, self-referential character, the content of the second-order [self-attributive] judgment is locked (self-referentially) onto the first-order content which it both contains and takes as its subject matter" (1988, 659–660; cf. Heil 1988; Gertler 2000, 2001; Heil and Gertler describe such thoughts as introspective while Burge appears not to think of self-knowledge so structured as introspective: 1998, 244; see also 1988, 652). In judging that I am thinking of a banana, I thereby necessarily think of a banana: The self-attributive judgment contains, as a part, the very thought self-ascribed, and thus cannot be false. In a 1996 essay, Burge extends his remarks to include not just self-attributive "thoughts" as targets but also (certain types of) "judgments" (e.g., "I judge, herewith, that there are physical entities" and other judgments with "herewith"-like reflexivity, 92)

Shoemaker (1994a, 1994b, 1994c) deploys the containment idea very differently, and over a much wider array of introspective targets, including conscious states like pains and propositional attitudes like belief. Shoemaker speculates that the relevant containment relation holds not between the *contents* or *concepts employed* in the target state and in the self-ascriptive state but rather between their neural realizations in the brain. To develop this point, Shoemaker distinguishes between a mental state's "core realization" and its "total realization". One might think of mental processes as transpiring in fairly narrow regions of the brain (their core realization), and yet, Shoemaker suggests, it's not as though we could simply carve off those regions from all others and still have the mental state in question. To be the mental state it is, the process must be embedded in a larger causal network involving more of the brain (the total realization). Relationships of containment or overlap between core realization and total realization between the target state and the self-ascriptive judgment might then underwrite introspective accuracy. For example, the total brain-state realization of the state of pain may simply be a subset of the total brain-state realization of the state of believing that one is in pain. Introspective accuracy might then be explained by the fact that the introspective judgment is not an independently existing state.

More recently, philosophers have applied Burge-like content-containment models (as opposed to Shoemaker-like realization-containment models) to self-knowledge of conscious states, or "phenomenology", in particular—for example, Gertler (2001), Papineau (2002), Chalmers (2003), and Horgan and Kriegel (2007). Husserl (1913/1982) offers an early phenomenal containment approach, arguing that we can at any time put our "cogitatio"—our conscious experiences—consciously before us through a kind of mental glancing, with the self-perception that arises containing as a part the conscious experience toward which it is directed, and incapable of existing without it. Papineau offers a "quotational" account on which in introspection we self-attribute "the experience: ___", where the blank is completed by the experience itself. Chalmers writes that "direct phenomenal beliefs" about our experiences are "partly *constituted* by an underlying phenomenal quality", in that the two will be tightly coupled across "a wide range of nearby conceptually possible cases" (2003, 235).

One possible difficulty with such accounts is that while it seems plausible to suppose that an introspective thought or judgment might contain another thought or judgment as a part, it's less clear how a self-attributive judgment or belief might contain a piece of conscious experience as a part. Beliefs, and other belief-like mental states like judgments, one might think, contain *concepts*, not conscious experiences, as their constituents (Fodor 1998); or, alternatively, one might think that beliefs are functional or dispositional patterns of response to input (Dennett 1987; Schwitzgebel 2002), again rendering it unclear how a piece of phenomenology could be part of belief. Perhaps with this concern in mind, advocates of containment accounts often appeal to "phenomenal concepts" that are, like the introspective judgments to which they contribute, partly constituted by the the conscious experiences that are the contents of those concepts. Such concepts are often thought to be obtained by demonstrative attention to our conscious experiences as they are ongoing.

It would seem, at least, that beliefs, concepts, or judgments containing pieces of phenomenology would have to expire once the phenomenology has passed and thus that the introspective judgments could not used in later inferences without recreating the state in question. Chalmers (2003) concedes the temporal locality of such phenomenology-containing introspective judgments and consequently their limited use in speech and in making generalizations. Papineau (2002), in contrast, embraces a theory in which the imaginative recreation of phenomenology in thinking about past experience is commonplace.

2.3.2 Self-Shaping

Although we can seemingly at least sometimes arrive at true self ascriptions through the self-shaping and the self-expression procedures (B and C) described at the beginning of Section 2.3, and although such procedures may meet the first three conditions on an account of introspection as described in Section 1.1—that is, they may (depending on how they are described and developed) be procedures that can yield only knowledge or judgments (or at least self-ascriptions) about one's own currently ongoing or very recently past mental states—few philosophers would describe such procedures as “introspective”. Nonetheless, they warrant brief treatment here, partly for the same reason self/other parity accounts warranted treatment in Section 2.1 above—that is, as skeptical accounts suggesting that the scope of introspection may be considerably narrower than is generally thought—and partly as background for the “transparency” accounts to be discussed in Section 2.3.4 below, with which they are often married.

It is difficult to find accounts of self-knowledge that stress the self-shaping technique in its purest, forward-looking, causal form—perhaps because it's clear that self-knowledge must involve considerably more than this (Gertler 2011). However, McGeer (1996, 2008; McGeer and Pettit 2002) puts considerable emphasis on self-shaping, writing that “we learn to use our intentional self-ascriptions to instill or reinforce tendencies and inclinations that fit with these ascriptions, even though such tendencies and inclinations may at best have been only nascent at the time we first made the judgments” (1996, 510). If I describe myself as brave in battle, or as a committed vegetarian —especially if I do so publicly—I create commitments and expectations for myself that help to make those self-ascriptions true. McGeer compares self-knowledge to the knowledge a driver has, as opposed to a passenger, of where the car is going: The driver, unlike the passenger, can make it the case that the car goes where she says it is going (505).

There are also strains in Dennett (though Dennett may not have an entirely consistent view on these matters; see Schwitzgebel 2007) that suggest either a self-fulfillment or a self-shaping view. In some places, Dennett compares “introspective” self-reports about consciousness to works of fiction, immune to refutation in the same way that fictional claims are—one could no more go wrong about one's consciousness, Dennett says, than Doyle could go wrong about the color of Holmes's easy chair (e.g., 1991, 81, 94). Such remarks are consistent with either an anti-realist view of fiction (there are no facts about the easy chair or about consciousness; see 366–367) or a self-fulfillment or self-shaping realist view (Doyle *creates* facts about Holmes as he thinks or writes about him; we create facts about what it's like to be us in thinking or making claims about our consciousness, as perhaps on 81 and 94). More moderately, in discussing attitudes, Dennett emphasizes how the act of formulating an attitude in language—for example, when ordering a menu item—can involve self-attributing a degree of specification in one's attitudes that was not present before, thereby committing one to, and partially or wholly creating, the specific attitude self-ascribed (1987, 20).

2.3.3 Expressivism

Wittgenstein writes:

> [H]ow does a human being learn the meaning of the names of sensations?—of the word “pain” for example. Here is one possibility: words are connected with the primitive, the natural, expressions of the sensation and used in their place. A child has hurt himself and he cries; and then adults talk to him and teach him exclamations and, later, sentences. They teach the child new pain-behaviour.
>
> “So you are saying that the word ‘pain’ really means crying?”—On the contrary: the verbal expression of pain replaces crying and does not describe it. (1953/1968, sec. 244)

And

> “It can't be said of me at all (except perhaps as a joke) that I *know* I am in pain. What is it supposed to mean—except perhaps that I *am* in pain?” (1953/1968, sec. 246).

On Wittgenstein's view, it is both true that I am in pain and that I say of myself that I am in pain, but the utterance in no way emerges from a process of *detecting* one's pain.

A simple expressivist view—sometimes attributed to Wittgenstein on the basis of these and related passages—denies that the expressive utterances (e.g., “that hurts!”) genuinely ascribe mental states to the individuals uttering them. Such a view faces serious difficulties accommodating the evident semantics of self-ascriptive utterances, including their use in inference and the apparent symmetries between present-tense and past-tense uses and between first-person and third-person uses (Wright 1998; Bar-On 2004). Consequently, Bar-On advocates, instead, what she calls a neo-expressivist view according to which expressive utterances can share logical and semantic structure with non-expressive utterances, despite the epistemic differences between them.

Expressivists have not always been clear about exactly the range of target mental states expressible in this way, but it seems plausible that at least in principle some true (or apt) self-ascriptions could arise in this manner, with no intervening introspective self-detection. The question would then be whether this is how we *generally* arrive at true self-ascriptions, for some particular class of mental states, or whether some more archetypically introspective process is also available. (For a more detailed treatment of expressivism, consult the section about the expressivist model of self-knowledge in the entry self-knowledge.)

2.3.4 Transparency

Evans writes:

> [I]n making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world. If someone asks me, “Do you think there is going to be a third world war?”, I must attend, in answering him, to precisely the same outward phenomena as I would attend to if I were answering the question “Will there be a third world war?” I get myself into the position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*. (1982, 225)

Transparency approaches to self-knowledge, like Evans', emphasize cases in which it seems that one arrives at an accurate self-ascription not by means of attending to, or thinking about, one's own mental states, but rather by means of attending to or thinking about the external states of the world that the target mental states are about. Note that this claim has both a negative and a positive aspect: We do *not* learn about our minds by as it were gazing inward; and we *do* learn about our minds by reflecting on the aspects of the world that our mental states are about. The positive and negative theses are separable: A pluralist might accept the positive thesis without the negative one; an advocate of a self/other parity theory or an expressivist account of self-knowledge (with respect to a certain class of target states) might accept the negative thesis without the positive. (N.B.: In the philosophical literature on self-knowledge “transparency” is also sometimes used to mean something like self-intimation in the sense of Section 4.1.1 below, for example in Wright 1998; Bilgrami 2006. This is a completely different usage, not to be confused with the present usage.) Because transparency accounts stress the outward focus of our thought in arriving at self-ascriptions, calling such accounts accounts of “introspection” strains against the etymology of the term. Nonetheless, some prominent advocates of transparency accounts, such as Dretske (1995) and Tye (2000), offer them explicitly as accounts of introspection.

The range of target states to which transparency applies is a matter of some dispute. Among philosophers who accept something like transparency, belief is generally regarded as transparent (Gordon 1995, 2007; Gallois 1996; Moran 2001; Fernández 2003; Byrne 2005). Perceptual states or perceptual experiences are also often regarded as transparent in the relevant sense. Harman's example is the most cited:

> When Eloise sees a tree before her, the colors she experiences are all experienced as features of the tree and its surroundings. None of them are experienced as intrinsic features of her experience. Nor does she experience any features of anything as intrinsic features of her experiences. And that is true of you too. There is nothing special about Eloise's visual experience. When you see a tree, you do not experience any features as intrinsic features of your experience. Look at a tree and try to turn your attention to intrinsic features of your visual experience. I predict you will find that the only features there to turn your attention to will be features of the presented tree. (Harman 1990, 667)

Harman's emphasis here is on the negative thesis, which goes back at least to Moore (1903; though Moore does not unambiguously endorse it). The view that it is impossible to attend directly to perceptual experience has recently been especially stressed by Tye (1995, 2000, 2002; see also Evans 1982; Van Gulick 1993; Shoemaker 1994a; Dretske 1995; Martin 2002; Stoljar 2004), and directly conflicts with accounts according to which we learn about our sensory experience primarily by directing introspective attention to it (e.g., Goldman 2006; Petitmengin 2006; Hill 2009; Siewert 2012; and back at least to Wundt 1888 and Titchener 1908/1973).

Gordon (2007) argues (contra Nichols and Stich 2003 and Goldman 2006) that Evans-like *ascent routines* (ascending from “*p*” to “I believe that *p*”) can drive the accurate self-ascription of all the attitudes, not just belief. He makes his case by wedding the transparency thesis to something like an expressive account of self-ascription: To answer a question about what I want—for example, which flavor ice cream do I want?—I think not about my desires but rather about the different flavors available, and then I *express* the resulting attitude self-ascriptively. Similarly for hopes, fears, wishes, intentions, regrets, etc. Gordon points out that from a very early age, before they likely have any self-ascriptive intent, children learn to express their attitudes self-ascriptively, for example with simple phrases like “[I] want banana!” (see also Bar-On 2004).

The transparency thesis is in fact consistent, not just with expressivism, but with any of the four non-detection-based self-ascription procedures described at the beginning of this section (and indeed Aydede and Güzeldere 2005 attempt to reconcile aspects of the transparency view with a broadly detection-like approach to introspection). This manifold compatibility highlights the fact that by itself the transparency thesis does not go far toward a positive view of the mechanisms of self-knowledge.

Moran (2001) brings together transparency and self-shaping in his *commissive* account of self-knowledge. Moran argues that normally when we are prompted to think about what we believe, desire, or intend (and he limits his account primarily to these three mental states), we reflect on the (outward) phenomena in question and make up our minds about what *to* believe, desire, or do. Rather than attempting to detect a pre-existing state, we open or re-open the matter and come to a resolution. Since we normally do believe, desire, and intend what we resolve to believe, desire, and do, we can therefore accurately self-ascribe those attitudes. Falvey (2000) embraces a similar view, and furthermore joins it with expressivism, a move Moran resists. (See also Falvey 2000; Boyle 2009; and see the discussion of the commitment model of self-knowledge in the entry self-knowledge for a more detailed discussion of commissive accounts.)

Byrne (2005) and Dretske (1995) bring together transparency and something like a derivational model of self-knowledge—a model on which I derive the conclusion that I believe that *P* directly from *P* itself, or the conclusion that I am representing *x* as *F* from the fact that *x* is *F*—a fact which must of course, to serve as a premise in the derivation, be represented (or believed) by me. Byrne argues that just as one might abide by the following epistemic rule:

> DOORBELL: If the doorbell rings, believe that there is someone at the door

so also might someone abide by the rule:

> BEL: If *P*, believe that you believe that *P*.

To determine whether you believe that *P*, first determine whether *P* is the case, then follow the rule BEL. Byrne (2011a, 2011b, 2011c, 2012) offers similar accounts of self-knowledge of intention, thinking, seeing, and desire.

Dretske analogizes introspection to ordinary cases of "displaced perception"—cases in which one perceives that something is the case by way of directly perceiving some other thing (e.g., hearing that the mail carrier has arrived by hearing the dog's barking; seeing that you weigh 110 pounds by seeing the dial on the bathroom scale): One perceives that one represents *x* as *F* by way of perceiving the *F*-ness of *x*. Dretske notes, however, two points of disanalogy between the cases. In the case of hearing that the mail carrier has arrived by hearing the dog's bark, the conclusion (that the mail carrier has arrived) is only established if the premise about the dog's barking is true, and furthermore it depends on a defeasible connecting belief, that the dog's barking is a reliable indicator of the mail's arrival. In the introspective case, however, the inference, if it is an inference, does not require the truth of the premise about *x*'s being *F*. Even if *x* is not *F*, the conclusion that I'm representing *x* as *F* is supported. Nor does there seem to be any sort of defeasible connecting belief.

Tye also emphasizes transparency in his account of introspection, though he limits his remarks to the introspection of conscious experience or "phenomenal character". In his 2000 book, Tye develops a view like Dretske's, analogizing introspection to displaced perception, though Tye unlike Dretske explicitly denies that inference is involved, instead proposing a mechanism similar to the sort of mechanism envisioned by simple monitoring accounts like those of Nichols and Stich (2003; see Section 2.2.1 above), a reliable process that, in the case of perceptual self-awareness, takes awareness of external things as its input and yields as its output awareness of phenomenal character. (The key difference between Tye's 2000 account on the one hand and the Nichols and Stich account on the other that warrants the classification of Tye's view here rather than in the section on self-detection models is this: Tye rejects the idea that the process is one of internal detection, while Nichols and Stich stress that idea. To adjudicate the dispute between those two positions, and to determine whether it might, in fact, be merely nominal, it would be helpful to have a clearer sense than has so far been given of what it means to say that one subpersonal system detects, or “monitors” or “scans”, the states or contents of another.) However, in his 2009 book, Tye rejects the displaced perception model in favor of a version of the transparency view that *identifies* phenomenal character with external qualities in the world, so that perceiving features of the world just is perceiving phenomenal character—a view that he recognizes is then charged with the difficult task of explaining how phenomenal character is a property (or “quality”) of external objects rather than, as is generally assumed, a property only of experiences of those objects.

Several authors have challenged the idea that sensory experience necessarily eludes attention—that is, they have denied the central claim of transparency theories about sensory experience. Block (1996), Kind (2003), and Smith (2008) have argued that phosphenes—those little lights you see when you press on your eyes—and visual blurriness are aspects of sensory experiences that can be directly attended. Siewert (2004) has argued that what's intuitively appealing in the transparency view is primarily the observation that in reflecting on sensory experience one does not *withdraw* attention from the objects sensed; but, he argues, this is compatible with also devoting a certain sort of attention to the sensory experience itself. In early discussions of attention, perceptual attention was sometimes distinguished from “intellectual attention” (James 1890/1981; Baldwin 1901–1905; see also Peacocke 1998; Mole 2011), that is, from the kind of attention we can devote to purely imagined word puzzles or to philosophical issues. If non-sensory forms of attention are possible, then the transparency thesis for sensory experience will require restatement: Is it only sensory attention to sensory experience that is impossible? Or is it any kind of attention whatsoever? Simply to say we don't attend sensorily to our mental states is to make only a modest claim, akin to the claim that we see objects rather than seeing our visual experiences of objects; but to say that we cannot attend to our mental states even intellectually appears extreme. In light of this, it remains unclear how to cast the transparency intuition to better bring out the core idea that is meant to be conveyed by the slogan that introspecting sensory experience is not a matter of attending to one's own mind.

2.4 Introspective Pluralism

Philosophers discussing self-knowledge often write as if approaches highlighting one of these methods of generating self-ascriptions conflict with approaches that highlight other of these methods, and also as if approaches of this general sort conflict with self-detection approaches (Section 2.2 above). While conflicts will certainly exist between different accounts intended to serve as *exhaustive* approaches to self-knowledge, it is implausible that any one or even any few of these approaches to self-knowledge is exhaustive. Plausibly, all of the non-self-detection approaches described above can lead, at least occasionally, to accurate self-ascriptions. Enthusiasts for another of the models, or for a self-detection model, needn't deny this. It also seems hard to deny that we at least *sometimes* reach conclusions about our mental lives based on the kind of theoretical inference or self-interpretation emphasized by advocates of self/other parity accounts (Section 2.1 above). Finally, even philosophers concerned about strong or oversimple self-scanning views might wish to grant that the mind can do *some* sort of tracking of its own present or recently past states—for example, when we trace back a stream of recently past thoughts that presumably can't (because past) be self-ascribed by self-fulfillment, self-shaping, self-expression, or transparency methods.

Schwitzgebel (2012) elevates this pluralism into a kind of negative account of introspection. Introspective judgments, he says, arise from a shifting confluence of many processes, recruited opportunistically, none of which can be called introspection proper. Just as there is no single, unified faculty of poster-taking-in that one employs when trying to take in a poster at a psychological conference or science fair, there is, on Schwitzgebel's view, no single, unified faculty of introspection or one underlying core process. Instead, the introspector, like the poster-viewer, brings to bear a diverse range of cognitive resources as suits the occasion. However, he says, the process wouldn't be worth calling "introspective" unless the introspector aimed to reach a judgment about her current or very recently past conscious experience, in a way that uses at least some resources specific to the first-person case, and in a way that involves some relatively direct sensitivity to the target state.

3. The Role of Introspection in Scientific Psychology

3.1 The Rise of Introspective Psychology as a Science

Philosophers have long made introspective claims about the human mind—or, to speak more cautiously, they've made claims seemingly at least in part introspectively grounded. Aristotle (3rd c. BCE/1961) asserts that thought does not occur without imagery. Mengzi (3rd c. BCE/2008) argues that our hearts are pleased by moral goodness and revolted by evil, even if the pleasure and revulsion are not evident in our outward behavior. Berkeley finds in himself no “abstract ideas” like that of a triangle that is, in Locke's terms “neither oblique, nor rectangle, neither equilateral, equicrural, nor scalenon, but all and none of these at once” (Berkeley 1710/1965, 12; Locke 1689/1975, 596). James Mill (1829/1878) attempts a catalog of the varieties of sense experience.

Although a number of early modern philosophers had aimed to initiate the scientific study of the mind, it wasn't until the middle of the 19th century—with the appearance of *quantitative introspective methods*, especially regarding sensory consciousness—that the study of the mind took shape as a progressive, mathematical, laboratory-based science. Early quantitative psychologists such as Helmholtz (1856/1962), Fechner (1860/1964), and Wundt (1896/1902) sought quantitative answers to questions like: By how much must two physical stimuli differ for the experiences of them to differ noticeably? How weak a stimulus can still be consciously perceived? What is the mathematical relationship between stimulus intensity and the intensity of the resulting sensation? (The Weber-Fechner law holds that the relationship is logarithmic.) Along what dimensions, exactly, can sense experience vary? (The “color solid” [see the link to the Munsell solid in Other Internet Resources, below], for example, characterizes color experience by appeal to just three dimensions of variation: hue, saturation, and lightness or brightness.) Although from very early on, psychologists also employed non-introspective methods (e.g., performance on memory tests, reaction times), most early characterizations of the field stood introspection at the center. James, for example, wrote that “introspective observation is what we have to rely on first and foremost and always” (1890/1981, 185).

In contrast with the dominant philosophical tradition that has, since Descartes, stressed the special privilege or at least high accuracy of introspective judgments about consciousness (see Section 4.1 below) many early introspective psychologists held that the introspection of currently ongoing or recently past conscious experience is difficult and prone to error if the introspective observer is insufficiently trained. Wundt, for example, reportedly did not credit the introspective reports of people with fewer than 50,000 trials of practice in observing their conscious experience (Boring 1953). Titchener, a leading American introspective psychologist, wrote a 1600-page introspective training manual for students, arguing that introspective observation is at least as difficult as observation in the physical sciences (Titchener 1901–1905; see also Wundt 1874/1908; Müller 1904; for contemporary discussions of introspective training see Varela 1996; Nahmias 2002; Schwitzgebel 2011b). This difference in optimism about untrained introspection may partly reflect differences in the types of judgments foregrounded in the two disciplines. Philosophers stressing privilege tend to focus on coarse and (seemingly) simple judgments such as "I'm having a visual experience of redness" or "I believe it's raining". The projects of interest to introspective psychologists often required much finer judgments—such as determining with mathematical precision whether one visual sensation has twice the "intensity" of another or determining along what dimensions emotional experience can vary.

3.2 Early Skepticism about Introspective Observation

Early introspective psychologists' theoretical discussions of the nature of introspection were often framed in reaction to skepticism about the scientific viability of introspection, especially the concern that the introspective act interferes with or destroys the mental state or process that is its target.[1]) The most influential formulation of this concern was Comte's:

> But as for observing in the same way *intellectual* phenomena at the time of their actual presence, that is a manifest impossibility. The thinker cannot divide himself into two, of whom one reasons whilst the other observes him reason. The organ observed and the organ observing being, in this case, identical, how could observation take place? This pretended psychological method is then radically null and void (1830, using the translation of James 1890/1981, 188).

Introspective psychologists tended to react to this concern in one of three ways. The most concessive approach—recommended, for example, by James (1890/1981; see also Mill 1865/1961; Lyons 1986)—was to grant Comte's point for *concurrent* introspection, that is, introspection simultaneous with the target state or process, and to emphasize in contrast *immediate retrospection*, that is, reflecting on or attending to the target process (usually a conscious experience) very shortly *after* it occurs. Since the scientific observation occurs only after the target process is complete, it does not interfere with that process; but of course the delay between the process and the observation must be as brief as possible to ensure that the process is accurately remembered.

Brentano (1874/1973) responded to Comte's concern by distinguishing between "inner observation" [*innere Beobachtung*] and "inner perception" [*innere Wahrnehmung*]. Observation, as Brentano characterizes it, involves dedicating full attention to a phenomenon, with the aim of apprehending it accurately. This dedication of attention necessarily interferes with the process to be observed if the process is a mental one; therefore, he says, inner observation is problematic as a scientific psychological method. Inner *perception*, in contrast, according to Brentano, does not involve attention to our mental lives and thus does not objectionably disturb them. While our "attention is turned toward a different object … we are able to perceive, incidentally, the mental processes which are directed toward that object" (1874/1973, 30). Brentano concedes that inner perception necessarily lacks the advantages of attentive observation, so he recommends conjoining it with retrospective methods.

Wundt (1888) agrees with Comte and Brentano that observation necessarily involves attention and so often interferes with the process to be observed, if that process is an inner, psychological one. To a much greater extent than Brentano, however, Wundt emphasizes the importance to scientific psychology of direct attention to experience, including planful and controlled variation. The psychological method of “inner perception” is, for Wundt, the method of holding and attentively manipulating a memory image or reproduction of a past psychological process. Although Wundt sees some value in this retrospective method, he thinks it has two crucial shortcomings: First, one can only work with what one remembers of the process in question—the manipulation of a memory-image cannot discover new elements. And second, foreign elements may be unintentionally introduced through association—one might confuse one's memory of a process with one's memory of another associated process or object.

Therefore, Wundt suggests, the science of psychology must depend upon the attentive observation of mental processes as they occur. He argues that those who think attention necessarily distorts the target mental process are too pessimistic. A *subclass* of mental processes remains relatively unperturbed by attentive observation—the “simpler” mental processes, especially of perception (1896/1902, 27–28). The experience of seeing red, Wundt claims, is more or less the same whether or not one is attending to the psychological fact that one is experiencing redness. Wundt also suggests that the basic processes of memory, feeling, and volition can be observed systematically and without excessive disruption. These *alone*, he thinks, can be studied by introspective psychology (see also Wundt 1874/1904; 1896/1902; 1907). Other aspects of our psychology must be approached through non-introspective methods such as the observation of language, mythology, culture, and human and animal development.

3.3 The Decline of Scientific Introspection

Although introspective psychologists were able to build scientific consensus on some issues concerning sense experience—issues such as the limits of sensory perception in various modalities and some of the contours of variation in sensory experience—by the early 20th century it was becoming clear that on many issues consensus was elusive. The most famous dispute concerned the existence of “imageless thought” (see the discussion of the imageless thought controversy in the entry mental imagery; see also Humphrey 1951; Kusch 1999); but other topics proved similarly resistant such as the structure of emotion or “feeling” (James 1890/1981; Külpe 1893/1895; Wundt 1896/1902; Titchener 1908/1973) and the experiential changes brought about by shifts in attention (Wundt 1896/1902; Pillsbury 1908; Titchener 1908/1973; Chapman 1933).

By the 1910s, behaviorism (which focused simply on the relationship between outward stimuli and behavioral response) had declared war on introspective psychology, portraying it as bogged down in irresolvable disputes between differing introspective “experts”, and also rebuking the introspectivists' passive taxonomizing of experience, recommending that psychology focus instead on socially useful paradigms for modifying behavior (e.g., Watson 1913). In the 1920s and 1930s, introspective studies were increasingly marginalized. Although strict behaviorism declined in the 1960s and 1970s, its main replacement, cognitivist functionalism (which treats functionally defined internal cognitive processes as central to psychological inquiry), generally continued to share behaviorism's disdain of introspective methods.

Psychophysics (the study of the relationship between physical sensory input and consequent psychological state or response), where the introspective psychologists had found their greatest success, underwent a subtle shift in this period from a focus on *subjective* methods—methods that involve asking subjects to report on their experiences or percepts—to a focus on *objective* methods such as asking subjects to report on states of the outside world, including insisting that subjects guess even when they feel they don't know or have no relevant conscious experience (especially with the rise of “signal detection theory” in psychophysics: Green and Swets 1966; Cheesman and Merikle 1986; Macmillan and Creelman 1991; Merikle, Smilek, and Eastwood 2001). Perhaps in accord with transparency views of introspection (Section 2.3.4 above), the two types of instruction to subjects seem very similar (compare the subjective “tell me if you visually experience a flash of light” with the objective “tell me if the light flashes”). On the other hand, perhaps in tension with transparency views, subjective and objective instructions seem sometimes to differ importantly, especially in cases of known illusion, Gestalt effects such as perceived grouping, stimuli near the limits of perceivability, and the experience of ambiguous figures (Boring 1921; Merikle, Smilek, and Eastwood 2001; Siewert 2004).

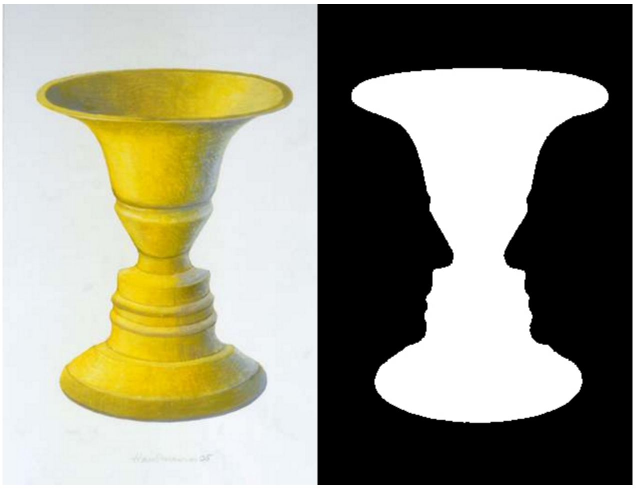
3.4 The Re-Emergence of Scientific Introspection

In no period, however, were introspective methods *entirely* abandoned by psychologists, and in the last few decades, they have begun to make something of a comeback, especially with the rise of the interdisciplinary field of “consciousness studies” (see, e.g., Jack and Roepstorff, eds., 2003, 2004). Ericsson and Simon (1984/1993; to be discussed further in Section 4.2.3 below) have advocated the use of “think-aloud protocols” and immediately retrospective reports in the study of problem solving. Other researchers have emphasized introspective methods in the study of imagery (Marks 1985; Kosslyn, Reisbert, and Behrmann 2006) and emotion (Lambie and Marcel 2002; Barrett et al. 2007).

Beeper methodologies have been developed to facilitate immediate retrospection, especially by Hurlburt (1990, 2011; Hurlburt and Heavey 2006; Hurlburt and Schwitzgebel 2007) and Csikszentmihalyi (Larson and Csikszentmihalyi 1983; Hektner, Schmidt, and Csikszentmihalyi 2007). Traditional immediately retrospective methods required the introspective observer in the laboratory somehow to intentionally refrain from introspecting the target experience as it occurs, arguably a difficult task. Hurlburt and Csikszentmihalyi, in contrast, give participants beepers to wear during ordinary, everyday activity. The beepers are timed to sound only at long intervals, surprising participants and triggering an immediately retrospective assessment of their “inner experience”, emotion, or thoughts in the moment before the beep.

Introspective or subjective reports of conscious experience have also played an important role in the search for the “neural correlates of consciousness” (as reviewed in Rees and Frith 2007; Tononi and Koch 2008; Prinz 2012; see also Varela 1996). One paradigm is for researchers to present ambiguous sensory stimuli, holding them constant over an extended period, noting what neural changes correlate with changes in subjective reports of experience. For example, in “binocular rivalry” methods, two different images (e.g., a face and a house) are presented, one to each eye. Participants typically say that only one image is visible at a time, with the visible image switching every few seconds.

Researchers have sometimes reported finding evidence that activity in "early" visual areas (such as V1) is not temporally coupled with reported changes in visual experience, while changes in conscious percept are better temporally coupled with activity in frontal and parietal areas further downstream and to large-scale changes in neural synchronization or oscillation; but the evidence is disputed (Lumer, Friston, and Rees 1998; Tong et al. 1998; Tononi et al. 1998; Polonsky et al. 2000; Kreiman, Fried, and Koch 2002; Moutoussis and Zeki 2002; Tong, Meng, and Blake 2006; Kamphuisen, Bauer, and van Ee 2008; Sandberg et al. 2013; Ishiku and Zeki 2014). Another version of the ambiguous sensory stimuli paradigm involves presenting the subject with an ambiguous figure such as the Rubin faces-vase figure:



Using this paradigm, researchers have found neuronal changes both in early visual areas and in later areas, as well as changes in widespread neuronal synchrony, that correspond temporally with subjective reports of flipping between one way and another of seeing the ambiguous figure (Kleinschmidt et al. 1998; Rodriguez et al. 1999; Ilg et al. 2008; Parkkonen et al. 2008; de Graaf et al. 2011). In masking paradigms, stimuli are briefly presented then followed by a "mask". On some trials, subjects report seeing the stimuli, while on others they don't. In trials in which the subject reports that stimulus was visually experienced, researchers have tended to find higher levels of activity through at least some of the downstream visual pathways as well as spontaneous electrical oscillations near 40 Hz (Dehaene et al. 2001; Summerfield, Jack, and Burgess 2002; Del Cul, Baillet, and Dehaene 2007; Quiroga et al. 2008). However, it remains contentious how properly to interpret such attempts to find neural correlates of consciousness (Noë and Thompson 2004; Overgaard, Sandberg, and Jensen 2008; Tononi and Koch 2008; Dehaene and Changeux 2011; Aru, Bachmann, Singer, and Melloni 2012; de Graaf, Hsieh, and Sack 2012).

If we report our attitudes by introspecting upon them, then much of survey research is also introspective, though psychologists have not generally explicitly described it as such. As with subjective vs. objective methods in psychophysics, there appears to be only a slight difference between subjectively phrased questions ("Do you approve of the President's handling of the war?", "Do you think gay marriage should be legalized?") and objectively phrased questions ("Has the President handled the war well?", "Should gay marriage be legalized?"). This would seem to support the observation at the core of transparency theory (discussed in Section 2.3.4 above) that questions about the mind and questions about the outside world often call for the same type of reflection.

4. The Accuracy of Introspection

4.1 Varieties of Privilege

It's plausible to suppose that people have some sort of *privileged access* to at least some of their own mental states or processes: You know about your own mind, or at least some aspects of it, in a different way and better than you know about other people's minds, and maybe also in a different way and better than you know about the outside world. Consider pain. It seems you know your own pains differently and better than you know mine, differently and (perhaps) better than you know about the coffee cup in your hand. If so, perhaps that special “first-person” privileged knowledge arises through something like introspection, in one or more of the senses described in Section 2 above.

Just as there is a diversity of methods for acquiring knowledge of or reaching judgments about one's own mental states and processes, to which the label “introspection” applies with more or less or disputable accuracy, so also is there a diversity of forms of “privileged access”, with different kinds of privilege and to which the idea of access applies with more or less or disputable accuracy. And as one might expect, the different introspective methods do not all align equally well with the different varieties of privilege.

4.1.1 Varieties of Perfection: Infallibility, Indubitability, Incorrigibility, and Self-Intimation

A substantial philosophical tradition, going back at least to Descartes (1637/1985; 1641/1984; also Augustine c. 420 C.E./1998), ascribes a kind of epistemic perfection to at least some of our judgments (or thoughts or beliefs or knowledge) about our own minds—infallibility, indubitability, incorrigibility, or self-intimation. Consider the judgment (thought, belief, etc.) that *P*, where *P* is a proposition self-ascribing a mental state or process (for example *P* might be *I am in pain*, or *I believe that it is snowing*, or *I am thinking of a dachshund*). The judgment that *P* is *infallible* just in case, if I make that judgment, it is not possible that *P* is false. It is *indubitable* just in case, if I make the judgment, it is not possible for me to doubt the truth of *P*. It is *incorrigible* just in case, if I make the judgment, it is not possible for anyone else to show that *P* is false. And it is *self-intimating* if it is not possible for *P* to be true without my reaching the judgment (thought, belief, etc.) that it is true. Note that the direction of implication for the last of these is the reverse of the first three. Infallibility, indubitability, and incorrigibility all have the form: “If I judge (think, believe, etc.) that *P*, then …”, while self-intimation has the form “If *P*, then I judge (think, believe, etc.) that *P*”. All four theses also admit of weakening by adding conditions to the antecedent “if” clause (e.g., “If I judge that *P* as a result of normal introspective processes, then …”). (See Alston 1971 for a helpful dissection of these distinctions; all admit of variations and nuance. Also note that some philosophers [e.g. Ayer 1936/1946; Armstrong 1963; Chalmers 2003; Tye 2009] use “incorrigibility” to mean infallibility as defined here, while others [e.g., Ayer 1963; Alston 1971; Rorty 1970; Dennett 2000] use it with the more etymologically specific meaning of [something like] “incapable of correction”.)

Descartes (1641/1984) famously endorsed the indubitability of “I think”, which he extends also to such mental states as doubting, understanding, affirming, and seeming to have sensory perceptions. He also appears to claim that the thought or affirmation that I am in such states is infallibly true. He was followed in this—especially in his infallibilism—by Locke (1690/1975), Hume (1739/1978), twentieth-century thinkers such as Husserl (1913/1982), Ayer (1936/1946, 1963), Lewis (1946), and the early Shoemaker (1963), and many others. Historical arguments for indubitability and infallibility have tended to center on intuitive appeals to the apparent impossibility of doubting or going wrong about such matters as whether one is having a thought with a certain content or is experiencing pain or having a visual experience as of seeing red.

Recent infallibilists have added to this intuitive appeal structural arguments based on self-fulfillment accounts of introspection or self-knowledge (see Section 2.3.1 above)—generally while also narrowing the scope of infallibility, for example to thoughts about thoughts (Burge 1988, 1996), or to "pure" phenomenal judgments about consciousness (Chalmers 2003; see also Wright 1998; Gertler 2001; Horgan, Tienson, and Graham 2006; Horgan and Kriegel 2007; Tye 2009; with important predecessors in Brentano 1874/1973; Husserl 1913/1982). The intuitive idea behind all these structural arguments is that somehow the self-ascriptive thought or judgment *contains* the mental state or process self-ascribed: the thought that I am thinking of a pink elephant contains the thought of a pink elephant; the judgment that I am having a visual experience of redness contains the red experience itself.

In contrast, self/other parity (Section 2.1) and self-detection (Section 2.2) accounts of introspection or self-knowledge appear to stand in tension with infallibilism. If introspection or self-knowledge involves a causal process from a mental state to an ontologically distinct self-ascription of that state, it appears that, however reliable such a process may generally be, there is inevitably room in principle for interference and error. Minimally, it seems, stroke, quantum accident, or clever neurosurgery could break otherwise generally reliable relationships between target mental states and the self-ascriptions of those states. Similar considerations apply to self-shaping (Section 2.3.2) and expressivist (Section 2.3.3) accounts, to the extent that these are interpreted causally rather than constitutively.

Introspective incorrigibility, as opposed to either infallibility or indubitability, was held by Rorty (1970) to be "the mark of the mental"—and thus as applying to a wide range of mental states—and has also been embraced more recently by Dennett (2000, 2002). The idea behind incorrigibility, recall, is that no one else could show your self-ascriptions to be false; or we might say, more qualifiedly and a bit differently, that if you arrive at the right kind of self-ascriptive judgment (perhaps an introspectively based judgment about a currently ongoing conscious process that survives critical reflection), then no one else, perhaps not even you in the future, aware of this, can rationally hold that judgment to be mistaken. If I judge that right now I am in severe pain, and I do so as a result of considering introspectively whether I am indeed in such pain (as opposed to, say, merely inferring that I am in pain based on outward behavior), and if I pause to think carefully about whether I really am in pain and conclude that I indeed am, then no one else who is aware of this can rationally believe that I'm not in pain, regardless of what my outward behavior might be (say, calm and relaxed) or what shows up in the course of brain imaging (say, no activation in brain centers normally associated with pain).

Incorrigibility does not imply infallibility: I may not actually be in pain, even if no one could *show* that I'm not. Consequently, incorrigibility is compatible with a broader array of sources of self-knowledge than is infallibility. Neither Rorty nor Dennett, for example, appear to defend incorrigibility by appeal to self-fulfillment accounts of introspection (though in both cases, interpreting their positive accounts is difficult). Causal accounts of self-knowledge may be compatible with incorrigibility if the causal connections underwriting the incorrigible judgments are vastly more trustworthy than judgments obtained without the benefit of this sort of privileged access. Of course, unless one embraces a strict self-fulfillment account, with its attendant infallibilism, one will want to rule out abnormal cases such as quantum accident; hence the need for qualifications.

Self-intimating mental states are those such that, if a person (or at least a person with the right background capacities) has them, she necessarily believes or judges or knows that she does. Conscious states are often held to be in some sense self-intimating, in that the mere having of them involves, requires, or implies some sort of representation or awareness of those states. Brentano argues that consciousness, for example, of an outward stimulus like a sound, “clearly occurs together with consciousness of this consciousness”, that is, the consciousness is “of the whole mental act in which the sound is presented and in which the consciousness itself exists concomitantly” (1874/1995, 129; see also phenomenological approaches to self-consciousness). Recent “higher order” and “same order” theories of consciousness (Armstrong 1968; Rosenthal 1990, 2005; Gennaro 1996; Lycan 1996; Carruthers 2005; Kriegel 2009; see also higher-order theories of consciousness) explain consciousness in terms of some thought, perception, or representation of the mental state that is conscious—the presence of that thought, perception, or representation being what makes the target state conscious. (On same order theories, the target mental state, or an aspect of it, represents itself, with no need for a distinct higher order state.) Thus, Horgan and others have described consciousness as “self-presenting” (Horgan, Tienson, and Graham 2005; Horgan and Kriegel 2007; the usage appears to follow Chisholm 1981, but Chisholm actually has an indubitability rather than a self-intimation thesis in mind). Shoemaker (1995) argues that beliefs—as long as they are “available” (i.e., readily deployed in inference, assent, practical reasoning, etc.), which needn't require that they are occurrently conscious—are self-intimating for individuals with sufficient cognitive capacity. Shoemaker's idea is that if the belief that *P* is available in the relevant sense, then one is disposed to do things like say “I believe *P*”, and such dispositions are themselves constitutive of believing that one believes that *P*.

Self-intimation claims (unlike infallibility, indubitability, and incorrigibility claims) are not usually cast as claims about “introspection”. This may be because knowledge acquired through self-intimation would appear to be constant and automatic, thus violating the effort condition on introspection (condition 6 in Section 1.1 above).

4.1.2 Weaker Guarantees

A number of philosophers have argued for forms of first-person privilege involving some sort of epistemic guarantee—not just conditional accuracy as a matter of empirical fact, but something more robust than that—without embracing infallibility, indubitability, incorrigibility, or self-intimation in the senses described in Section 4.1.1 above.

Shoemaker (1968), for example, argues that self-knowledge of certain psychological facts such as “I am waving my arm” or “I see a canary”, when arrived at “in the ordinary way (without the aid of mirrors, etc.)”, is *immune to error through misidentification relative to the first-person pronoun* (see also Campbell 1999; Pryor 1999; Bar-On 2004; Hamilton 2008). That is, although one may be wrong about waving one's arm (perhaps the nerves to your arm were recently severed unbeknownst to you) or about seeing a canary (perhaps it's a goldfinch), one cannot be wrong due to mistakenly identifying the person waving the arm or seeing the canary *as you*, when in fact it is someone else. This immunity arises, Shoemaker argues, because there is no need for identification in the first place, and thus no opportunity for *mis*-identification. In this respect, Shoemaker argues, knowledge that a particular arm that is moving is your arm (not immune to misidentification since maybe it's someone else's arm, misidentified in the mirror) is different from the knowledge that *you* are moving your arm—knowledge, that is, of what Searle (1983) calls an "intention in action".

Shoemaker has also argued for the conceptual impossibility of introspective *self-blindness* with respect to one's beliefs, desires, and intentions, and for somewhat different reasons one's pains (1988, 1994b). A self-blind creature, by Shoemaker's definition, would be a rational creature with a conception of the relevant mental states, and who can entertain the thought that she has this or that belief, desire, intention, or pain, but who nonetheless utterly lacks introspective access to the type of mental state in question. A self-blind creature could still gain "third person" knowledge of the mental states in question, through observing her own behavior, reading textbooks, and the like. (Thus, strict self/other parity accounts of self-knowledge of the sort described in Section 2.1 are accounts according to which one is self-blind in Shoemaker's sense.) Shoemaker's case against self-blindness with respect to belief turns on the dilemma of whether the self-blind creature can avoid "Moore-paradoxical" sentences (see Moore 1942, 1944/1993; Shoemaker 1995) like "it's raining but I don't believe that it's raining" in which the subject asserts both *P* and that she doesn't believe that *P*. If the subject is truly self-blind, Shoemaker suggests, there should be cases in which her best evidence is both that *P* and that she doesn't believe that *P* (the latter, perhaps, based on misleading facts about her behavior). But if the subject asserts "*P* but I don't believe that *P*" in such cases, she does not (contra the initial supposition) really have a rational command of the nature of belief and assertion; and thus it's not a genuine case of self-blindness as originally intended. Alternatively, perhaps the creature can reliably avoid such Moore-paradoxical sentences, self-attributing belief in an apparently normal way. But then, Shoemaker suggests, it seems that she is indistinguishable from normal people in thought and behavior and hence not self-blind. For desire, intention, and pain, too, Shoemaker aims to reveal incoherences between having a rational command of the concepts in question and behaving as though one were systematically ignorant of or mistaken about those states. Shoemaker uses his case against self-blindness as part of his argument against self-detection accounts of introspection (described in Section 2.2 above): If introspection were a matter of detecting the presence of states that exist independently of the introspective judgment or belief, then it ought to be possible for the faculty enabling the detection to break down entirely, as in the case of blindness, deafness, etc., in outward perception (see also Nichols and Stich 2003, who argue that schizophrenia provides such a case).

Burge has influentially asserted that *brute errors* about "present, ordinary, accessible propositional attitudes [such as belief and desire]" are impossible or at least subject to "severe limits"—where a "brute error" is an error that "indicates no rational failure and no malfunction in the mistaken individual" such as commonly occur in ordinary perception due to "misleading natural conditions or look-alike substitutes" (1988, 657–658; 1996, 103–104). However, Burge offers little argument for this claim, apart from the argument mentioned in Sections 2.3.1 and 4.1.1 above that for certain sorts of self-ascriptions error in general (and not just "brute error") is impossible, due to the "self-verifying" nature of such self-ascriptions.

Dretske (1995, 2004) argues that we have infallible knowledge of the *content* of our attitudes without necessarily knowing (or even having a very good idea about) the *attitude* we take toward those contents. For example, if I believe that *it will rain tomorrow*, I have infallibly accurate information, which I may then access introspectively, regarding the presence of a mental state with a certain content—the content "it will rain tomorrow"—but I may often have little or no information about the fact that my attitude toward that content is the particular attitude it is—belief, in this case (as opposed to supposition or hope). This view follows from Dretske's accepting something like a containment account of the introspection of the content of the attitude (the introspective judgment employing the same content as the target attitude; see Section 2.3.1 above, especially the discussion of Burge), while he sees knowledge of the attitude one has toward that content as requiring complex information about the causal role and history of that mental state.

Transcendental arguments for the accuracy of certain sorts of self-knowledge offer a different sort of epistemic guarantee—"transcendental arguments" being arguments that assume the existence of some sort of experience or capacity, then develop insights about the background conditions necessary for that experience or capacity, and finally conclude that those background conditions must in fact be met. Burge (1996; see also Shoemaker 1988) argues that to be capable of "critical reasoning" one must be able to recognize one's own attitudes, knowledgeably evaluating, identifying, and reviewing one's beliefs, desires, commitments, suppositions, etc., where these mental states are known to be the states they are. Since we are (by assumption, for the sake of transcendental argument) capable of critical reasoning, we must have some knowledge of our attitudes. Bilgrami (2006) argues that we can only be held responsible for actions if we know the beliefs and desires that "rationalize" our actions; since we can (by assumption) sometimes be held responsible, we must sometimes know our beliefs and desires. Wright (1989) argues that the "language game" of ascribing "intentional states" such as belief and desire to oneself and others requires as a background condition that self-ascriptions have special authority within that game. Given that we successfully play this language game, we must indeed have the special authority that we assume and others grant us in the context of the game.

4.1.3 Privilege Without Guarantee

Developing an analogy from Wright (1998), if it's your turn with the kaleidoscope, you have a type of privileged perspective on the shapes and colors it presents. If someone else in the room wants to know what color dominates, for example, the most straightforward course would be to ask you. But this type of privileged access comes with no guarantee. At least in principle, you might be quite wrong about the tumbling shapes. You might be dazzled by afterimages, or momentarily confused, or hallucinating, or (unbeknownst to you) colorblind. (Yes, people often don't know they are colorblind, a point stressed by Kornblith 1998.) It is also at least in principle possible that others may know better than you, perhaps even systematically so, what is transpiring in the kaleidoscope. You might think the figure shows octagonal symmetry, but the rest of us, familiar with the kaleidoscope's design, might know that the symmetry is hexagonal. A brilliant engineer may invent a kaleidoscope state detector that can dependably reveal from outside the shape, color, and position of the tumbling chunks.

Wright raises this analogy to suggest that people's privilege with respect to certain aspects of their mental lives must be different from that of the person with the kaleidoscope; but other philosophers, especially those who embrace self-detection accounts of introspection, should find the analogy at least somewhat apt: Introspective privilege is akin to the privilege of having a unique and advantageous sensory perspective on something. Metaphorically speaking, we are the only ones who can gaze directly at our attitudes or our stream of experience, while others must rely on us or on outward signs.

Less metaphorically, in generating introspective judgments (or beliefs or knowledge) about one's own mentality one employs a detection process available to no one else. It is then an empirical question how accurate the deliverances of this process are; but on the assumption that the deliverances are in a broad range of conditions at least somewhat accurate and more accurate than the typical judgments other people make about those same aspects of your mind, you have a “privileged” perspective. Typically, advocates of self-detection models of introspection regard the mechanism or cognitive process generating introspective judgments or beliefs as highly reliable in roughly this way, but not infallible, and not immune to correction by other people (Armstrong 1968; Churchland 1988; Hill 1981, 2009; Lycan 1996; Nichols and Stich 2003; Goldman 2000, 2006).

4.2 Empirical Evidence on the Accuracy of Introspection

The arguments of the previous section are a priori in at least the broad sense of that term (the psychologists' sense): They depend on general conceptual considerations and armchair folk psychology rather than on empirical research. To these might be added the argument, due to Boghossian (1989) that “externalism” about the content of our attitudes (the view that our attitudes depend constitutively not just on what is going on internally but also on facts about our environment; Putnam 1975; Burge 1979) seems to problematize introspective self-knowledge of those attitudes. This issue will not be treated here, since it is amply covered in the entries on externalism about mental content and externalism and self-knowledge.

Now we turn to empirical research on our self-knowledge of those aspects of our minds often thought to be accessible to introspection. Since character traits are not generally regarded as introspectible aspects of our mentality, we'll skip the large literature on the accuracy or inaccuracy of our judgments about them (e.g., Taylor and Brown 1988; Paulhus and John 1998; Funder 1999; Vazire 2010; see also Haybron's 2008 skeptical perspective on our knowledge of how happy we are); nor will we discuss self-knowledge of subpersonal, nonconscious mental processes, such as the processes underlying visual recognition of color and shape.

As a general matter, while a priori accounts of the epistemology of introspection have tended to stress its privilege and accuracy, empirical accounts have tended to stress its failures.

4.2.1 Of the Causes of Attitudes and Behavior

Perhaps the most famous argument in the psychological literature on introspection and self-knowledge is Nisbett and Wilson's argument that we have remarkably poor knowledge of the causes of, and processes underlying, our behavior and attitudes (Nisbett and Wilson 1977; Nisbett and Ross 1980; Wilson 2002). Section 2.1 above briefly mentioned their emblematic finding that people in a shopping mall were often ignorant of a major factor—position—influencing their judgments about the quality of pairs of stockings. In Nisbett and Bellows (1977), also briefly mentioned above, participants were asked to assess the influence of various factors on their judgments about features of a supposed job applicant. As in Nisbett and Wilson's stocking study, participants denied the influence of some factors that were in fact influential; for example, they denied that the information that they would meet the applicant influenced their judgments about the applicant's flexibility. (It actually had a major influence, as assessed by comparing the judgments of participants who were told and not told that they would meet the applicant.) Participants also attributed influence to factors that were not in fact influential; for example, they falsely reported that the information that the applicant accidentally knocked over a cup of coffee during the interview influenced “how sympathetic the person seems” to them. Nisbett and Bellows found that ordinary observers' hypothetical ratings of the influence of the various factors on the various judgments closely paralleled the participants' own ratings of the factors influencing them—a finding used by Nisbett to argue that people have no special access to causal influences on their judgments and instead rely on the same sorts of theoretical considerations outside observers rely on (the self/other parity view described in Section 2.1). Despite some objections (such as White 1988), both psychologists and philosophers now tend to accept Nisbett's and Wilson's view that there is at best only a modest first-person advantage in assessing the factors influencing our judgments and behavior.

In series of experiments, Gazzaniga (1995) presented commissurotomy patients (people with severed corpus callosum) with different visual stimuli to each hemisphere of the brain. With cross-hemispheric communication severely impaired due to the commissurotomy, the left hemisphere, controlling speech, had information about one part of the visual stimulus, while the right hemisphere, controlling some aspects of movement (especially the left hand) had information about a different part. Gazzaniga reported finding that when these “split brain” patients were asked to explain why they did something, when that action was clearly caused by input to the right, non-verbal hemisphere, the left hemisphere would sometimes fluently confabulate an explanation. For example, Gazzaniga reports presenting an instruction like “laugh” to the right hemisphere, making the patient laugh. When asked why he laughed, the patient would say something like “You guys come up and test us every month. What a way to make a living!” (1393). When a chicken claw was shown to the left hemisphere and snow scene to the right, and the patient was asked to select an appropriate picture from an array, the right hand would point to a chicken and the left hand to a snow shovel, and when asked why she selected those two things, the patient would say something like “Oh, that's simple. The chicken claw goes with the chicken and you need a shovel to clean out the chicken shed” (ibid.). Similar confabulation about motives is sometimes (but not always) seen in people whose behavior is, unbeknownst to them, driven by post-hypnotic suggestion (Richet 1884; Moll 1889/1911), and in disorders such as hemineglect (anosognosia), blindness denial (Anton's syndrome), and Korsakoff's syndrome (Hirstein 2005).

In a normal population, Johansson and collaborators (Johansson et al. 2005; Johansson et al. 2006) manually displayed to participants pairs of pictures of women's faces. On each trial, the participant was to point to the face he found more attractive. The picture of that face was then centered before the participant while the other face was hidden. On some trials, participants were asked to explain the reasons for their choices while continuing to look at the selected face. On a few key trials, the experimenters used sleight-of-hand to present to the participant the face that was *not* selected as though it had been the face selected. Strikingly, the switch was noticed only 28% of the time. What's more, when the change was not detected participants actually gave explanations for their choice that appealed to specific features of the unselected face that were not possessed by the selected face 13% of the time. For example, one participant claimed to have chosen the face before him “because I love blondes” when in fact he had chosen a dark-haired face (Johansson et al. 2006, 690). Johansson and colleagues failed to find any systematic differences in the explanations of choice between the manipulated and non-manipulated trials, using a wide variety of measures. They found, for example, no difference in linguistic markers of confidence (including pauses in speech), emotionality, specificity of detail, complexity or length of description, or general position in semantic space. These results, like Nisbett's and Wilson's, suggest that at least some of the time when people think they are explaining the bases of their decisions, they are instead merely theorizing or confabulating.

Wegner has found that people can often be manipulated into believing that they willed or intended behavior that is in fact caused by another person's manipulation and, conversely, that they exerted no control over movements that were in fact their own—as with Ouija boards, with or without a cheating, intentionally directive confederate (Wegner and Wheatley 1999; Wegner 2002). The literature on “cognitive dissonance” is replete with cases in which participants' attitudes appear to change for reasons they do, or would, deny. According to cognitive dissonance theory, when people behave or appear to behave counternormatively (e.g., incompetently, foolishly, immorally), they will tend to adjust their attitudes so as to make the behavior seem less counternormative or “dissonant” (Festinger 1957; Aronson 1968; Cooper and Fazio 1984; Stone and Cooper 2001). For example, people induced to falsely describe as enjoyable a monotonous task they've just completed will tend, later, to report having a more positive attitude toward the task then those not induced to lie (though much less so if they were handsomely paid to lie in which case the behavior is not clearly counternormative; Festinger and Carlsmith 1959; but see Bem 1967, 1972 for an argument that the attitude doesn't change but only the report of it). Presumably, if such attitude changes were known to the subject they would generally fail to have their dissonance-reducing effect. Research psychologists have also confirmed such familiar phenomena as “sour grapes” (Lyubomirsky and Ross 1999; Kay, Jiminez, and Jost 2002) and “self-deception” (Mele 2001) which presumably also involve ignorance of the factors driving the relevant judgments and actions. And of course the Freudian psychoanalytic tradition has also long held that people often have only poor knowledge of their motives and the influences on their attitudes (Wollheim 1981; Cavell 2006).

In light of this empirical research, no major philosopher now holds (perhaps no major philosopher ever held) that we have infallible, indubitable, incorrigible, or self-intimating knowledge of the causes of our judgments, decisions, and behavior. Perhaps weaker forms of privilege also come under threat. But the question arises: Whatever failures there may be in assessing the causes of our attitudes and behavior, are those failures failures of *introspection*, properly construed? Psychologists tend to cast these results as failures of “introspection”, but if it turns out that a very different and more trustworthy process underwrites our knowledge of some other aspects of our minds—such as what our present attitudes are (however caused) or our currently ongoing or recently past conscious experience—then perhaps we can call only *that* process introspection, thereby retaining some robust form of introspective privilege while acceding to the psychological consensus regarding (what we would now call non-introspective) first-person knowledge of causes. Indeed, few contemporary philosophical accounts of introspection or privileged self-knowledge highlight, as the primary locus of privilege, the causes of our attitudes and behavior (though Bilgrami 2006 is a notable exception). Thus, the literature reviewed in this section can be interpreted as suggesting that the causes of our behavior are not, after all, the sorts of things to which we have introspective access.

4.2.2 Of Attitudes

Research psychologists have generally not been as skeptical of our knowledge of our attitudes as they have been of our knowledge of the causes of our attitudes (Section 4.2.1 above). In fact, many of the same experiments that purport to show inaccurate knowledge of the causes of our attitudes nonetheless rely unguardedly on self-report for assessment of the attitudes themselves—a feature of those experiments criticized by Bem (1967). Attitudinal surveys in psychology and social science generally rely on participants' self-report as the principal source of evidence about attitudes (de Vaus 1985/2002; Sirken et al. (eds.) 1999). However, as in the case of motives and causes, there's a long tradition in clinical psychology skeptical of our self-knowledge of our attitudes, giving a large role to “unconscious” motives and attitudes.

A key challenge in assessing the accuracy of people's beliefs or judgments about their attitudes is the difficulty of accurately measuring attitudes independently of self-report. There is at present no tractable measure of attitude that is generally seen by philosophers as overriding individuals' own reports about their attitudes. However, in the psychological literature, “implicit” measures of attitudes —measures of attitudes that do not reply on self-report—have recently been gaining considerable attention (see Wittenbrink and Schwarz, eds., 2007; Petty, Fazio, and Briñol, eds., 2009). Such measures are sometimes thought capable of revealing unconscious attitudes or implicit attitudes either unavailable to introspection or erroneously introspected (Wilson, Lindsey, and Schooler 2000; Kihlstrom 2004; Lane et al. 2007; though see Hahn et al. forthcoming).

Much of the leading research on implicit attitude measures has concerned racism, in accord with the view that racist attitudes, though common, are considered socially undesirable and therefore often not self-ascribed even when present. For example, Campbell, Kruskal, and Wallace (1966) explored the use of seating distance as an index of racial attitudes, noting that racially Black and White students tended to aggregate in classroom seating arrangements. Using facial electromyography (EMG), Vanman et al. (1997) found (racially) White participants to display facial responses indicative of negative affect more frequently when asked to imagine co-operative activity with Black than with White partners—results interpreted as indicative of racist attitudes. Cunningham et al. (2004) showed White and Black faces to White participants while participants were undergoing fMRI brain imaging. They found less amygdala activation when participants looked at faces from their own group than when participants looked at other faces; and since amygdala activation is generally associated with negative emotion, they interpreted this tendency suggesting a negative attitude toward outgroup members (see also Hart et al 1990; and for discussion Ito and Cacioppo 2007).

Much of the recent implicit attitude research has focused on response priming and interference in speeded tasks. In priming research, a stimulus (the “prime”) is briefly displayed, followed by a mask that hides it, and then a second stimulus (the “target”) is displayed. The participant's task is to respond as swiftly as possible to the target, typically with a classification judgment. In *evaluative priming,* for example, the participant is primed with a positively or negatively valenced word or picture (e.g., snake), then asked to make a swift judgment about whether the subsequently presented target word (e.g., “disgusting”) is good or bad, or has some other feature (e.g., belongs to a particular category). Generally, negative primes will speed response for negative targets while delaying response for positive targets, and positive primes will do the reverse. Researchers have found that photographs of Black faces, whether presented visibly or whether presented so quickly as to be subliminal, tend to facilitate the categorization of negative targets and delay the categorization of positive targets for White participants—a result widely interpreted as revealing racist attitudes (Fazio et al. 1995; Dovidio et al. 1997; Wittenbrink, Judd, and Park 1997). In the *Implicit Association Test*, respondents are asked to respond disjunctively to combined categories, giving for example one response if they see either a dark-skinned face or a positively valenced word and a different response if they see either a light-skinned face or a negatively valenced word. As in evaluative priming tasks, White respondents tend to respond more slowly when asked to pair dark-skinned faces with positively valenced words than with negatively valenced words, which is interpreted as revealing a negative attitude or association (Greenwald, McGhee, and Schwartz 1998; Lane et al. 2007).

As mentioned above, such implicit measures are often interpreted as revealing attitudes to which people have poor or no introspective access. The evidence that people lack introspective knowledge of such attitudes generally turns on the low correlations between such implicit measures of racism and more explicit measures such as self-report—though due to the recognized social undesirability of racial prejudice, it is difficult to disentangle self-presentational from self-knowledge factors in self-reports (Fazio et al. 1995; Greenwald, McGhee, and Schwartz 1998; Wilson, Lindsey, and Schooler 2000; Greenwald and Nosek 2009). People who appear racist by implicit measures might disavow racism and inhibit racist patterns of response on explicit measures (such as when asked to rate the attractiveness of faces of different races) because they don't want to be *seen* as racist—a motivation that might drive them whether or not they have accurate self-knowledge of their racist attitudes. Still, it seems prima facie plausible that people have at best limited knowledge of the patterns of association that drive their responses on priming and other implicit measures.

But what do such tests really measure? In philosophy, Zimmerman (2007) and Gendler (2008a, 2008b) have argued that measures like the Implicit Association Test do not measure actual racist beliefs but rather something else, something under less rational control (Gendler calls them “aliefs”). In psychology, Gawronski and Bodenhausen (2006) advance a model according to which there is a substantial difference between implicit attitudes, defined in terms of associative processes, and explicit attitudes which have a propositional structure and are guided by standards of truth and consistency (see also Wilson, Lindsey, and Schooler 2000; Greenwald and Nosek 2009). On such a model, as on Zimmerman's and Gendler's views, a person with implicit racist associations may nonetheless have fully and genuinely egalitarian propositional beliefs. To the extent attitudes are held to be reflected in, or even defined by, our explicit judgments about the matter in question and also, differently but perhaps not wholly separably (see Section 2.3.4 above), our explicit judgments about our *attitudes* toward the matter in question, our self-knowledge would seem to be correspondingly secure and implicit measures beside the point. To the extent attitudes are held to crucially involve swift and automatic, or unreflective, patterns of reaction and association, our self-knowledge of them would appear to be correspondingly problematic, corrigible by data from implicit measures (Bohner and Dickel 2011; Schwitzgebel 2011a).

In a different vein, Carruthers (2011; see also Rosenthal 2001; Bem 1967, 1972) argues that the evidence of Nisbett, Gazzaniga, Wegner, and others (reviewed in Section 4.2.1 above) shows that people confabulate not just in reporting the *causes* of their attitudes but also in reporting the attitudes themselves. For example, Carruthers suggests that if someone in Nisbett and Wilson's famous 1977 study confabulates “I thought this pair was softest” as an explanation of her choice of the rightmost pair of stockings, she errs not only about the cause of her choice but also in ascribing to herself the judgment that the pair was softest. On this basis, Carruthers adopts a self/other parity view (see Section 2.1 above) of our self-knowledge of our attitudes, holding that we can only introspect, in the strict sense, conscious experiences like those that arise in perception and imagery.

4.2.3 Of Conscious Experience

Currently ongoing conscious experience—or maybe immediately past conscious experience (if we hold that introspective judgment must temporally follow the state or process introspected, or if we take seriously the concerns raised in Section 3.2 about the self-undermining of the introspective process)—is both the most universally acknowledged target of the introspective process and the target most commonly thought to be known with a high degree of privilege. Infallibility, indubitability, incorrigibility, and self-intimation claims (see Section 4.1.1) are most commonly made for self-knowledge of states such as being in pain or having a visual experience as of the color red, where these states are construed as qualitative states, or subjective experiences, or aspects of our phenomenology or consciousness. (All these terms are intended interchangeably to refer to what Block [1995], Chalmers [1996], and other contemporary philosophers call “phenomenal consciousness”.) If attitudes are sometimes conscious, then we might also be capable of introspecting those attitudes as part of our capacity to introspect conscious experience generally (Goldman 2006; Hill 2009).

It's difficult to study the accuracy of self-ascriptions of conscious experience for the same reasons it's difficult to study the accuracy of our self-ascriptions of attitudes (Section 4.2.2): There's no widely accepted measure to trump or confirm self-report. In the medical literature on pain, for example, no behavioral or physiological measure of pain is generally thought capable of overriding self-report of current pain, despite the fact that scaling issues remain a problem both within and especially between subjects (Williams, Davies, and Chadury 2000) as does retrospective assessment (Redelmeier and Kahneman 1996). When physiological markers of pain and self-report dissociate, it's by no means clear that the physiological marker should be taken as the more accurate index (for methodological recommendations see Price and Aydede 2005). Corresponding remarks apply to the case of pleasure (Haybron 2008).

As mentioned in Section 3.3 above, early introspective psychologists both asserted the difficulty of accurately introspecting conscious experience and achieved only mixed success in their attempts to obtain scientifically replicable (and thus presumably accurate) data through the use of trained introspectors. In some domains they achieved considerable success and replicability, such as in the construction of the “color solid” (a representation of the three primary dimensions of variation in color experience: hue, saturation, and lightness or brightness), the mapping of the size of “just noticeable differences” between sensations and the “liminal” threshold below which a stimulus is too faint to be experienced, and the (at least roughly) logarithmic relationship between the intensity of a sensory stimulus and the intensity of the resulting experience (the “Weber-Fechner law”). Contemporary psychophysics—the study of the relation between physical stimuli and the resulting sense experiences or percepts—is rooted in these early introspective studies. However, other sorts of phenomena proved resistant to cross-laboratory introspective consensus—such as the possibility or not of imageless thought (see the entry on “mental imagery”), the structure of emotion, and the experiential aspects of of attention. Perhaps these facts about the range of early introspective agreement and apparently intractable disagreement cast light on the range over which careful and well-trained introspection is and is not reliable.

Ericsson and Simon (1984/1993; Ericsson 2003) discuss and review relationships between the subject's performance on various problem-solving tasks, her concurrent verbalizations of conscious thoughts (“think aloud protocols”), and her immediately retrospective verbalizations. The existence of good relationships in the predicted directions in many problem-solving tasks lends empirical support to the view that people's reports about their stream of thoughts often accurately reflect those thoughts. For example, Ericsson and Simon find that think-aloud and retrospective reports of thought processes correlate with predicted patterns of eye movement and response latency. Ericsson and Simon also cite studies like that of Hamilton and Sanford (1978), who asked subjects to make yes or no judgments about whether pairs of letters were in alphabetical order (like MO) or not (like RP) and then to describe retrospectively their method for arriving at the judgments. When subjects retrospectively reported knowing the answer “automatically” without an intervening conscious process, reaction times were swift and did not depend on the distance between the letters. When subjects retrospectively reported “running through” a sequential series of letters (such as “LMNO” when prompted with “MO”) reaction times correlated nicely with reported length of run-through. On the other hand, Flavell, Green, and Flavell (1995) report gross and widespread introspective error about recently past and even current (conscious) thought in young children; and Smallwood and Schooler (2006) review literature that suggests that people are not especially good at detecting when their mind is wandering.

In the 20th century, philosophers arguing against infallibilism often devised hypothetical examples in which they suggested it was plausible to attribute introspective error; but even if such examples succeed, they are generally confined to far-fetched scenarios, pathological cases, or very minor or very brief mistakes (e.g., Armstrong 1963; Churchland 1988; Kornblith 1998, with an eye to the distinction between mistakes about current conscious experience and other sorts of mistakes). In the 21st century, philosophical critics of the accuracy of introspective judgments about consciousness shifted their focus to cases of widespread disagreement or (putative) error, either among ordinary people or among research specialists. Dennett (1991), Blackmore (2002), and Schwitzgebel (2011b), for example, argue that most people are badly mistaken about the nature of the experience of peripheral vision. These authors argue that people experience visual clarity only in a small and rapidly moving region of about 1–2 degrees of visual arc, contrary to the (they say) widespread impression most people have that they experience a substantially broader range of stable clarity in the visual field. Other recent arguments against the accuracy of introspective judgments about conscious experience turn on citing the widespread disagreement about whether there is a “phenomenology of thinking” beyond that of imagery and emotion, about whether sensory experience as a whole is “rich” (including for example constant tactile experience of one's feet in one's shoes) or “thin” (limited mostly just to what is in attention at any one time), and about the nature of visual imagery experience (Hurlburt and Schwitzgebel 2007; Bayne and Spener 2010; Schwitzgebel 2011b; though see Hohwy 2011). Irvine (2013) has argued that the methodological problems in this area are so severe that the term “consciousness” should be eliminated from scientific discourse as impossible to effectively operationalize or measure.

Bibliography

Alston, William P., 1971, “Varieties of privileged access”, *American Philosophical Quarterly*, 8: 223–241.
Amedi, Amir, Rafael Malach, and Alvaro Pascual-Leone, 2005, “Negative BOLD differentiates visual imagery and perception”, *Neuron*, 48: 859–872.
Aristotle, 3rd c. BCE/1961, *De Anima*, W.D. Ross (ed.), Oxford: Oxford University Press.
Armstrong, David M., 1963, “Is introspective knowledge incorrigible?”, *Philosophical Review*, 72:

417–432.

–––, 1968, *A materialist theory of the mind*, London: Routledge.

–––, 1981, *The nature of mind and other essays*, Ithaca, NY: Cornell University Press.

–––, 1999, *The mind-body problem*, Boulder, CO: Westview.

Aronson, Elliot, 1968, “Dissonance theory: Progress and problems”, in *Theories of cognitive consistency*, Robert P. Abelson, et al. (eds.), Chicago: Rand McNally, 112–139.

Aru, Jaan, Talis Bachmann, Wolf Singer, and Lucia Melloni, 2012, “Distilling the neural correlates of consciousness”, *Neuroscience and Biobehavioral Reviews*, 36: 737–746.

Augustinus, Aurelius, c. 420 C.E./1998, *The city of God against the pagans*, R.W. Dyson (tr.), Cambridge: Cambridge University Press.

Aydede, Murat, and Güven Güzeldere, 2005, “Cognitive architecture, concepts, and introspection: An information-theoretic solution to the problem of phenomenal consciousness”, *Noûs*, 39: 197–255.

Ayer, A.J., 1936/1946, *Language, truth, and logic*, 2nd ed., London: Gollancz.

–––, 1963, *The concept of a person*, New York: St. Martin's.

Baldwin, James Mark, 1901–1905, *Dictionary of philosophy and psychology*, New York: Macmillan.

Bar-On, Dorit, 2004, *Speaking my mind*, Oxford: Oxford.

Barrett, Lisa Feldman, Batja Mesquita, Kevin N. Ochsner, and James J. Gross, 2007, “The experience of emotion”, *Annual Review of Psychology*, 58: 373–403.

Bayne, Tim, and Michelle Montague, eds., 2011, *Cognitive phenomenology*, Oxford: Oxford University Press.

Bayne, Tim, and Maja Spener, 2010, “Introspective humility”, *Philosophical Issues*, 20, 1–22.

Bem, Daryl J., 1967, “Self-perception: An alternative interpretation of cognitive dissonance phenomena”, *Psychological Review*, 74: 183–200.

–––, 1972, “Self-perception theory”, *Advances in Experimental Social Psychology*, 6: 1–62.

Berkeley, George, 1710/1965, *A Treatise Concerning the Principles of Human Knowledge*, in *Principles, Dialogues, and Philosophical Correspondence*, Colin M. Turbayne (ed.), New York: Macmillan, 3–101.

Bilgrami, Akeel, 2006, *Self-knowledge and resentment*, Cambridge, MA: Harvard University Press.

Blackmore, Susan, 2002, “There is no stream of consciousness”, *Journal of Consciousness Studies*, 9(5–6), 17–28.

Block, Ned, 1995, “On a confusion about a function of consciousness”, *Behavioral and Brain Sciences*, 18: 227–247.

–––, 1996, “Mental paint and mental latex”, *Philosophical Issues*, 7: 19–49.

Boghossian, Paul, 1989, “Content and self-knowledge”, *Philosophical Topics*, 17: 5–26.

Bohner, Gerd, and Nina Dickel, 2011, “Attitudes and attitude change”, *Annual Review of Psychology*, 62: 391–417.

Boring, Edwin G., 1921, “The stimulus-error”, *American Journal of Psychology*, 32: 449–471.

–––, 1953, “A history of introspection”, *Psychological Bulletin*, 50: 169–189.

Boyle, Matthew, 2009, “Two kinds of self-knowledge”, *Philosophy & Phenomenological Research*, 78: 133–164.

Brentano, Franz, 1874/1995, *Psychology from an empirical standpoint*, 2nd English edition, Antos C. Rancurello, D. B. Terrell and Linda L. McAlister (trans.), New York: Routledge.

Burge, Tyler, 1979, “Individualism and the mental”, *Midwest Studies in Philosophy*, 4: 73–121.

–––, 1988, “Individualism and self-knowledge”, *Journal of Philosophy*, 85: 649–663.

–––, 1996, “Our entitlement to self-knowledge”, *Proceedings of the Aristotelian Society*, 96: 91–116.

–––, 1998, “Reason and the first person”, in *Knowing our own minds*, Crispin Wright, Barry C. Smith, and Cynthia Macdonald (eds.), Oxford: Oxford University Press, 243–270.

Byrne, Alex, 2005, “Introspection”, *Philosophical Topics*, 33(1): 79–104.

–––, 2011a, Knowing that I am thinking, in *Self-knowledge*, Anthony Hatzimoysis (ed.), Oxford: Oxford.

–––, 2011b, Knowing what I want, in *Consciousness and the Self*, JeeLoo Liu and John Perry (eds.), Cambridge: Cambridge.

–––, 2011c, Transparency, belief, intention, *Aristotelian Society Supplementary Volume,*, 85, 201-221.

–––, 2012, in *Introspection and Consciousness*, Declan Smithies and Daniel Stoljar (eds.), Oxford: Oxford.

Campbell, Donald T., William H. Kruskal, and William P. Wallace, 1966, “Seating aggregation as an index of attitude”, *Sociometry*, 29: 1–15.

Campbell, John, 1999, “Immunity to error through misidentification and the meaning of a referring term”, *Philosophical Topics*, 25(1–2), 89–104.

Carruthers, Peter, 2005, *Consciousness: Essays from a higher-order perspective*, Oxford: Oxford University Press.

–––, 2011, *The Opacity of Mind*, Oxford: Oxford University Press.

Cavell, Marcia, 2006, *Becoming a subject*, Oxford: Oxford University Press.

Chalmers, David J., 1996, *The conscious mind*, New York: Oxford.

–––, 2003, “The content and epistemology of phenomenal belief”, in *Consciousness: New philosophical perspectives*, Quentin Smith and Aleksandar Jokic (eds.), Oxford: Oxford, 220–272.

Chapman, Dwight W., 1933, “Attensity, clearness, and attention”, *American Journal of Psychology*, 45: 156–165.

Cheesman, Jim, and Philip M. Merikle, 1986, “Distinguishing conscious from unconscious perceptual processes”, *Canadian Journal of Psychology*, 40: 343–367.

Chisholm, Roderick M., 1981, *The first person*, Brighton, UK: Harvester.

Churchland, Paul M., 1988, *Matter and consciousness*, rev. ed., Cambridge, MA: MIT Press.

Comte, Auguste, 1830, *Cours de philosophie positive*, volume 1, Paris: Bacheleier, Libraire pour les Mathématiques.

Cooper, Joel, and Russell H. Fazio, 1984, “A new look at dissonance theory”, *Advances in Experimental Social Psychology*, 17: 229–266.

Cui, Xu, Cameron B. Jeter, Dongni Yang, P. Read Montague, and David M. Eagleman, 2007, “Vividness of mental imagery: Individual variability can be measured objectively”, *Vision Research*, 47: 474–478.

Cunningham, William A., et al., 2004, “Separable neural components in the processing of Black and White faces”, *Psychological Science*, 15: 806–813.

de Graaf, Tom A., Maartje c. de Jong, Rainer Goebel, Raymond van Ee, and Alexander T. Sack, 2011, “On the functional relevance of frontal cortex for passive and volunatarily controlled bistable vision”, *Cerebral Cortex*, 21: 2322–2331.

de Graaf, Tom A., Po-Jang Hsieh, and Alexander T. Sack, 2012, “The 'correlates' in neural correlates of consciousness”, *Neuroscience and Biobehavioral Reviews*, 36: 191–197.

De Vaus, David, 1985/2002, *Surveys in social research*, London: Routledge.

Dehaene, Stanislaus, et al., 2001, “Cerebral mechanisms of word masking and unconscious repetition priming”, *Nature Neuroscience*, 4: 752–758.

Dehaene, Stanislaus, and Jean-Pierre Changeux, 2011, “Experimental and theoretical approaches to conscious processing”, *Neuron*, 70: 200–227.

Del Cul, Antoine, Sylvain Baillet, and Stanislas Dehaene, 2007, “Brain dynamics underlying the nonlinear threshold for access to consciousness”, *PLoS Biology*, 5(10): e260).

Dennett, Daniel C., 1987, *The intentional stance*, Cambridge, MA: MIT Press.

–––, 1991, *Consciousness explained*, Boston: Little, Brown, and Co.

–––, 2000, “The case for rorts”, in *Rorty and his critics*, R.B. Brandom (ed.), Malden, MA: Blackwell, 91–101.

–––, 2002, “How could I be wrong? How wrong could I be?”, *Journal of Consciousness Studies*, 9(5–6): 13–6.

Descartes, René, 1637/1985, *Discourse on the method*, in *The philosophical writings of Descartes*, vol. 1, John Cottingham, Robert Stoothoff, and Dugald Murdoch (editors and translators), Cambridge: Cambridge University Press, 111–151.

–––, 1641/1984, *Meditations on first philosophy*, in *The philosophical writings of Descartes*, vol. 2, John Cottingham, Robert Stoothoff, and Dugald Murdoch (editors and translators,), Cambridge: Cambridge University Press, 1–62.

Dovidio, John F., Kerry Kawakami, Craig Johnson, Brenda Johnson, and Adaiah Howard, 1997, “On the nature of prejudice: Automatic and controlled processes”, *Journal of Experimental Social Psychology*, 33: 510–540.

Dretske, Fred, 1995, *Naturalizing the mind*, Cambridge, MA: MIT.

–––, 2004, “Knowing what you think vs. knowing that you think it”, in *The externalist challenge*, Richard Schantz (ed.), Berlin: Walter de Gruyter, 389–399.

Ebbinghaus, Hermann, 1885/1913, *Memory: A contribution to experimental psychology*, Henry A. Ruger and Clara E. Bussenius (translators), New York: Columbia.

Ericsson, K. Anders, 2003, “Valid and non-reactive verbalization of thoughts during performance of tasks: Towards a solution to the central problems of introspection as a source of scientific data”, *Journal of Consciousness Studies*, 10(9–10): 1–18.

Ericsson, K. Anders, and Herbert A. Simon, 1984/1993, *Protocol analysis*, rev. ed., Cambridge, MA: MIT.

Evans, Gareth, 1982, *The varieties of reference*, John McDowell (ed.), Oxford: Clarendon; New York: Oxford University Press.

Fazio, Russell H., Joni R. Jackson, Bridget C. Dunton, and Carol J. Williams, 1995, “Variability in automatic activation as an unobtrusive measure of racial attitudes: A bona fide pipeline?”, *Journal of Personality and Social Psychology*, 69(6): 1013–1027.

Fechner, Gustav, 1860/1964, *Elements of Psychophysics*, Helmut E. Adler, Davis H. Howes, and Edwin G. Boring (ed. and trans.), New York: Holt, Rinehart, and Winston.

Fernández, Jorgi, 2003, “Privileged access naturalized”, *Philosophical Quarterly*, 53: 352–372.

Festinger, Leon, 1957, *A theory of cognitive dissonance*, Stanford, CA: Stanford.

Festinger, Leon, and James M. Carlsmith, 1959, “Cognitive consequences of forced compliance”, *Journal of Abnormal and Social Psychology*, 58: 203–210.

Flavell, John H., Frances L. Green, and Eleanor R. Flavell, 1995, *Young children's knowledge about thinking*, *Monographs of the Society for Research in Child Development*, 60(1).

Fodor, Jerry A., 1983, *Modularity of mind*, Cambridge, MA: MIT.

–––, 1998, *Concepts: Where cognitive science went wrong*, Oxford: Oxford University Press.

Funder, David C., 1999, *Personality judgment*, London: Academic.

Gallois, Andre, 1996, *The world without, the mind within*, Cambridge: Cambridge.

Galton, Francis, 1869/1891, *Hereditary genius*, rev. ed., New York: Appleton.

Gardner, Sebastian, 1993, *Irrationality and the philosophy of psychoanalysis*, Cambridge: Cambridge University Press.

Gazzaniga, Michael S., 1995, “Consciousness and the cerebral hemispheres”, in *The Cognitive Neurosciences*, Michael S. Gazzaniga (ed.), Cambridge, MA: MIT, 1391–1400.

Gawronski, Bertram, and Galen V. Bodenhausen, 2006, “Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change”, *Psychological Bulletin*, 132: 692–731.

Gendler, Tamar Szabó, 2008a, “Alief and Belief”, *Journal of Philosophy*, 105: 634–663.

–––, 2008b, “Alief in Action, and Reaction”, *Mind & Language*, 23: 552–585.

Gennaro, Rocco J., 1996, *Consciousness and Self-Consciousness*, Amsterdam: John Benjamins.

Gertler, Brie, 2000, “The mechanics of self-knowledge”, *Philosophical Topics*, 28: 125–146.

–––, 2001, “Introspecting phenomenal states”, *Philosophy and Phenomenological Research*, 63: 305–328.

–––, 2011, “Self-knowledge and the transparency of belief”, in *Self-knowledge*, Anthony Hatzimoysis (ed.), Oxford: Oxford University Press.

Goldman, Alvin I., 1989, “Interpretation psychologized”, *Mind and Language*, 4: 161–185.

–––, 2000, “Can science know when you're conscious?”, *Journal of Consciousness Studies*, 7(5): 3–22.

–––, 2006, *Simulating minds*, Oxford: Oxford.

Gopnik, Alison, 1993a, “How we know our minds: The illusion of first-person knowledge of intentionality”, *Behavioral and Brain Sciences*, 16: 1–14.

–––, 1993b, “Psychopsychology”, *Consciousness and Cognition*, 2: 264–280.

Gopnik, Alison, and Andrew N. Meltzoff, 1994, “Minds, bodies and persons: Young children's understanding of the self and others as reflected in imitation and ‘theory of mind’ research”, in *Self-awareness in animals and humans*, Sue Taylor Parker, Robert W. Mitchell, and Maria L. Boccia (eds.), New York: Cambridge, 166–186.

Gordon, Robert M., 1995, “Simulation without introspection or inference from me to you”, in *Mental simulation*, Martin Davies and Tony Stone (eds.), Oxford: Blackwell.

–––, 2007, “Ascent routines for propositional attitudes”, *Synthese*, 159: 151–165.

Green, David M., and John A. Swets, 1966, *Signal detection theory and psychophysics*, Oxford: Wiley.

Greenwald, Anthony G., Debbie E. McGhee, and Jordan L.K. Schwartz, 1998, “Measuring individual differences in implicit cognition: The Implicit Association Test”, *Journal of Personality and Social Psychology*, 74: 1464–1480.

Greenwald, Anthony G., and Brian A. Nosek, 2009, “Attitudinal dissociation: What does it mean?”, in *Attitudes: Insights from the New Implicit Measures*, Richard E. Petty, Russell H. Fazio, and Pablo Briñol (eds.), New York: Taylor and Francis, 65–82.

Hahn, Adam, Charles M. Judd, Holen K. Hirsh, and Irene V. Blair, forthcoming, “Awareness of implicit attitudes”, *Journal of Experimental Psychology: General*.

Hamilton, Andy, 2007, “Memory and self-consciousness: Immunity to error through misidentification”, *Synthese*, 171: 409–417.

Hamilton, J.M.E., and A.J. Sanford, 1978, “The symbolic distance effect for alphabetic order judgements: A subjective report and reaction time analysis”, *Quarterly Journal of Experimental*

Psychology, 30: 33–43.

Harman, Gilbert, 1990, “The intrinsic quality of experience”, in James Tomberlin, (ed.), *Philosophical Perspectives*, 4, Atascadero, CA: Ridgeview, 31–52.

Hart, Allen J., Paul J. Whalen, Lisa M. Shin, Sean C. McInerney, Hakan Fischer, and Scott L. Rauch, 2000, “Differential response in the human amygdala to racial outgroup *vs* ingroup face stimuli”, *NeuroReport*, 11: 2351–2355.

Haybron, Daniel M., 2008, *The pursuit of unhappiness*, Oxford: Oxford University Press.

Hektner, Joel M., Jennifer A. Schmidt, and Mihaly Csikszentmihalyi, 2007, *Experience sampling method*, Thousand Oaks, CA: Sage.

Heil, John, 1988, “Privileged access”, *Mind*, 97: 238–251.

Helmholtz, Hermann, 1856/1962, *Helmholtz's Treatise on Physiological Optics*, James P.C. Southall (ed.), New York: Dover. [Translation based on 1924 edition.]

Hill, Christopher S., 1991, *Sensations: A defense of type materialism*, Cambridge: Cambridge University Press.

–––, 2009, *Consciousness*, Cambridge: Cambridge University Press.

Hirstein, William, 2005, *Brain fiction*, Cambridge, MA: MIT.

Hohwy, Jakob, 2011, “Phenomenal variability and introspective reliability”, *Mind & Language*, 26: 261–286.

Horgan, Terence, John L. Tienson, and George Graham, 2006, “Internal-world skepticism and mental self-presentation”, in *Self-representational approaches to consciousness*, Uriah Kriegel and Kenneth Williford (eds.), Cambridge, MA: MIT, 191–207.

Horgan, Terence, and Uriah Kriegel, 2007, “Phenomenal epistemology: What is consciousness that we may know it so well?”, *Philosophical Issues*, 17(1): 123–144.

Hume, David, 1739/1978, *A treatise of human nature*, L.A. Selby-Bigge and P.H. Nidditch (eds.), Oxford: Clarendon.

–––, 1748/1975, *An enquiry concerning human understanding*, in *Enquiries concerning human understanding and concerning the principles of morals*, L.A. Selby-Bigge and P.H. Nidditch (eds.), Oxford: Clarendon, 1–165.

Humphrey, George, 1951, *Thinking: An introduction to its experimental psychology*, London: Methuen.

Hurlburt, Russell T., 1990, *Sampling normal and schizophrenic inner experience*, New York: Plenum.

Hurlburt, Russell T., 2011, *Investigating pristine inner experience*, Cambridge: Cambridge.

Hurlburt, Russell T., and Christopher L. Heavey, 2006, *Exploring inner experience*, Amsterdam: John Benjamins.

Hurlburt, Russell T., and Eric Schwitzgebel, 2007, *Describing inner experience? Proponent meets skeptic*, Cambridge, MA: MIT.

Husserl, Edmund, 1913/1982, *Ideas*, Book I, T.E. Klein and W.E. Pohl (trs.), Dordrecht: Kluwer.

Ilg, Rüdiger, Afra M. Wohlschläger, Stefan Burazanis, Andreas Wöller, Sabine Nunnemann, and Mark Mühlau, 2008, “Neural correlates of spontaneous percept switches in ambiguous stimuli: An event-related functional magnetic resonance imaging study”, *European Journal of Neuroscience*, 28: 2325–2332.

Irvine, Elizabeth, 2013, *Consciousness as a scientific concept*, Dordrecht: Springer.

Ito, Tiffany A., and John T. Cacioppo, 2007, “Attitudes as mental and neural states of readiness”, in *Implicit measures of attitudes*, Bernd Wittenbrink and Norbert Schwarz (eds.), New York: Guilford, 125–158.

Jack, Anthony, and Andreas Roepstorff, 2003, *Trusting the subject*, vol. 1, special issue of the *Journal of Consciousness Studies*, no. 10(9–10).

–––, 2004, *Trusting the subject*, vol. 2, special issue of the *Journal of Consciousness Studies*, 11(7–8).

James, William, 1890/1981, *The principles of psychology*, Cambridge, MA: Harvard.

Jaynes, Julian, 1976, *The origin of consciousness in the breakdown of the bicameral mind*, New York: Houghton Mifflin.

Johansson, Petter, Lars Hall, Sverker Sikström, and Andreas Olsson, 2005, "Failure to detect mismatches between intention and outcome in a simple decision task", *Science*, 310: 116–119.

Johansson, Petter, Lars Hall, Sverker Sikström, Betty Tärning, and Andreas Lind, 2006, "How something can be said about telling more than we can know: On choice blindness and introspection", *Consciousness and Cognition*, 15: 673–692.

Kamphuisen, Allard, Markus Bauer, and Raymond van Ee, 2008, "No evidence for widespread synchronized networks in binocular rivalry: MEG frequency tagging entrains primary early visual cortex", *Journal of Vision*, 8(5): article 4.

Kant, Immanuel, 1781/1997, *The critique of pure reason*, Paul Guyer and Allen W. Wood (eds. and trs.), Cambridge: Cambridge.

Kay, Aaron C., Maria C. Jimenez, and John T. Jost, 2002, "Sour grapes, sweet lemons, and the anticipatory rationalization of the status quo", *Personality and Social Psychology Bulletin*, 28: 1300–1312.

Kihlstrom, John F., "Implicit methods in social psychology", in *The SAGE handbook of methods in social psychology*, Carol Sansone, Carolyn C. Morf, and A.T. Panter (eds.), Thousand Oaks, CA: Sage, 195–212.

Kind, Amy, 2003, "What's so transparent about transparency?", *Philosophical Studies*, 115: 225–244.

Kleinschmidt, A., C. Büchel, S. Zeki, and R.S.J. Frackowiak, 1998, "Human brain activity during spontaneously reversing perception of ambiguous figures", *Proceedings of the Royal Society B*, 265: 2427–2433.

Knapen, Tomas, Jan Brascamp, Joel Pearson, Raymond van Ee, and Randolph Blake, 2011, "The role of frontal and parietal areas in bistable perception", *Journal of Neuroscience*, 31: 10293–10301.

Kornblith, Hilary, 1998, "What is it like to be me?", *Australasian Journal of Philosophy*, 76: 48–60.

Kosslyn, Stephen M., Daniel Reisberg, and Marlene Behrmann, 2006, "Introspection and mechanism in mental imagery", in *The Dalai Lama at MIT*, Anne Harrington and Arthur Zajonc (eds.), Cambridge, MA: Harvard, 79–90.

Kreiman, Gabriel, Itzhak Fried, and Christof Koch, 2002, "Single-neuron correlates of subjective vision in the human medial temporal lobe", *Proceedings of the National Academy of Sciences*, 99: 8378–8383.

Kriegel, Uriah, 2009, *Subjective consciousness*, Oxford: Oxford.

Külpe, Oswald, 1893/1895, *Outlines of psychology*, London: Swan Sonnenschein.

Kusch, Martin, 1999, *Psychological knowledge*, London, Routledge.

Lambie, John A., and Anthony J. Marcel, 2002, "Consciousness and the varieties of emotion experience: A theoretical framework", *Psychological Review*, 109: 219–259.

Lane, Kristin A., Mahzarin R. Banaji, Brian A. Nosek, and Anthony G. Greenwald, 2007, "Understanding and using the Implicit Association Test: IV", in *Implicit measures of attitudes*, Bernd Wittenbrink and Norbert Schwarz (eds.), New York: Guilford, 59–102.

Larson, Reed, and Mihaly Csikszentmihalyi, 1983, "The Experience Sampling Method" in Harry T. Reis, (ed.), *Naturalistic approaches to studying social interaction*, San Francisco: Jossey-Bass, 41-56.

Lear, Jonathan, 1998, *Open-minded*, Cambridge, MA: Harvard.

Lewis, C.I., 1946, *An analysis of knowledge and valuation*, La Salle, IL: Open Court.

Locke, John, 1690/1975, *An essay concerning human understanding*, Peter H. Nidditch (ed.), Oxford: Oxford University Press.

Lumer, Erik D., Karl J. Friston, and Geraint Rees, 1998, "Neural correlates of perceptual rivalry in the human brain", *Science*, 280: 1930–1934.

Lycan, William G., 1996, *Consciousness and experience*, Cambridge, MA: MIT.

Lyons, William, 1986, *The disappearance of introspection*, Cambridge, MA: MIT.

Lyubomirsky, Sonja, and Lee Ross, 1999, "Changes in attractiveness of elected, rejected, and precluded alternatives: A comparison of happy and unhappy individuals", *Journal of Personality and Social Psychology*, 76: 988–1007.

Macmillan, Neil A., and C. Douglas Creelman, 1991, *Detection theory*, Cambridge: Cambridge University Press.

Marks, David F., 1985, "Imagery paradigms and methodology" *Journal of Mental Imagery*, 9: 93–105.

Marr, David, 1983, *Vision*, New York: Freeman.

Martin, Michael G.F., 2002, "The transparency of experience", *Mind and Language*, 17: 376–425.

Maudsley, Henry, 1867/1977, *Physiology and pathology of the mind*, Daniel N. Robinson (ed.), Washington, DC: University Publications of America.

McGeer, Victoria, 1996, "Is 'self-knowledge' an empirical problem? Renegotiating the space of philosophical explanation", *Journal of Philosophy*, 93: 483–515.

–––, 2008, "The moral development of first-person authority", *European Journal of Philosophy*, 16: 81–108.

McGeer, Victoria, and Philip Pettit, 2002, "The self-regulating mind", *Language and Communication*, 22: 281–299.

Mele, Alfred, 2001, *Self-deception unmasked*, Princeton, NJ: Princeton.

Mengzi, 3rd c. BCE/2008, B.W. Van Norden (tr.), Indianapolis: Hackett.

Merickle, Philip M., Daniel Smilek, and John D. Eastwood, 2001, "Perception without awareness: Perspectives from cognitive psychology", *Cognition*, 79: 115–134.

Mill, James, 1829/1878, *Analysis of the Phenomena of the Human Mind*, John Stuart Mill (ed.), London: Longmans, Green, Reader, and Dyer.

Mill, John Stuart, 1865/1961, *Auguste Comte and positivism*, Ann Arbor, MI: University of Michigan.

Mole, Christoper, 2011, *Attention is cognitive unison*, Oxford: Oxford University Press.

Moll, Albert, 1889/1911, *Hypnotism*, Arthur F. Hopkirk (ed.), New York: Charles Scribner's Sons.

Moore, George Edward, 1903, "The refutation of idealism", *Mind*, 12: 433–453.

–––, 1942, "A reply to my critics", in *The philosophy of G.E. Moore*, in P.A. Schilpp (ed.), New York: Tudor, 535–677.

–––, 1944/1993, "Moore's paradox", in G.E. Moore, *Selected writings*, Thomas Baldwin (ed.), London: Routledge, 207–212.

Moran, Richard, 2001, *Authority and estrangement*, Princeton: Princeton.

Müller, G.E., 1904, *Die Gesichtspunkte und die Tatsachen der psychophysischen Methodik*, Wiesbaden: J.F. Bergmann.

Nahmias, Eddy, 2002, "Verbal reports on the contents of consciousness: Reconsidering introspectionist methodology", *Psyche*, 8(21).

Nichols, Shaun, and Stephen P. Stich, 2003, *Mindreading*, Oxford: Oxford University Press.

Nisbett, Richard E., and Nancy Bellows, 1977, “Verbal reports about causal influences on social judgments: Private access versus public theories”, *Journal of Personality and Social Psychology*, 35: 613–624.

Nisbett, Richard E., and Lee Ross, 1980, *Human inference*, Englewood Cliffs, NJ: Prentice-Hall.

Nisbett, Richard E., and Timothy DeCamp Wilson, 1977, “Telling more than we can know: Verbal reports on mental processes”, *Psychological Review*, 84: 231–259.

Noë, Alva, 2004, *Action in perception*, Cambridge, MA: MIT Press.

Noë, Alva, and Evan Thompson, 2004, “Are there neural correlates of consciousness?”, *Journal of Consciousness Studies*, 11: (1): 3–28.

Overgaard, Morten, Kristian Sandberg, and Mads Jensen, 2008, “The neural correlate of consciousness?”, *Journal of Theoretical Biology*, 254: 713–715.

Papineau, David, 2002, *Thinking about consciousness*, Oxford: Oxford University Press.

Parkkonen, Lauri, Jesper Andersson, Matti Hämäläinen, and Riitta Hari, 2008, “Early visual brain areas reflect the percept of an ambiguous scene”, *Proceedings of the National Academy of Sciences*, 105: 20500–20504.

Paulhus, Delroy L., and Oliver P. John, 1998, “Egoistic and moralistic biases in self-perception: The interplay of self-deceptive styles with basic traits and motives”, *Journal of Personality*, 66: 1025–1060.

Peacocke, Christopher, 1998, “Conscious attitudes, attention, and self-knowledge”, in *Knowing our own minds*, Crispin Wright, Barry C. Smith, and Cynthia Macdonald (eds.), Oxford: Oxford University Press, 63–99.

Petitmengin, Claire, 2006, “Describing one's subjective experience in the second person: An interview method for the science of consciousness”, *Phenomenology and the Cognitive Sciences*, 5: 229–269.

Petty, Richard E., Russell H. Fazio, and Pablo Briñol (eds.), 2009, *Attitudes: Insights from the new implicit measures*, New York: Taylor and Francis.

Pillsbury, W.B., 1908, *Attention*, London: Swan Sonnenschein.

Price, Donald D., and Murat Aydede, 2005, “The experimental use of introspection in the scientific study of pain and its integration with third-person methodologies: The experiential-phenomenological approach”, in *Pain: New essays on its nature and the methodology of its study*,Murat Aydede (ed.), Cambridge, MA: MIT, 243–273.

Prinz, Jesse, 2004, “The fractionation of introspection”, *Journal of Consciousness Studies*, 11(7–8): 40–57.

–––, 2007, “Mental pointing: Phenomenal knowledge without concepts”, *Journal of Consciousness Studies*, 14(9–10): 184–211.

–––, 2012, *The conscious brain*, Oxford: Oxford.

Pryor, James, 1999, “Immunity to error through misidentification”, *Philosophical Topics*, 26(1–2): 271–304.

Putnam, Hilary, 1975, “The meaning of ‘meaning’” in Hilary Putnam, *Philosophical papers*, vol. 2, Cambridge: Cambridge University Press, 215–271.

Quiroga, R. Quian, R. Mukamel, E.A. Isham, and I. Fried, 2008, “Human single-neuron responses at the threshold of conscious recognition”, *Proceedings of the National Academy of Sciences*, 105: 3599–3604.

Redelmeier, Donald A., and Daniel Kahneman, 1996, “Patients' memories of painful medical treatments: Real-time and retrospective evaluations of two minimally invasive procedures", *Pain*, 66: 3–8.

Rees, Geraint, and Chris Frith, 2007, "Methodologies for identifying the neural correlates of consciousness", in *The Blackwell Companion to Consciousness*, Max Velmans and Susan Schneider (eds.), Malden, MA: Blackwell, 553–566.

Richet, Charles, 1884, *L'homme et l'intelligence*, Paris: F. Alcan.

Rodriguez, Eugenio, Nathalie George, Jean-Philippe Lachauz, Jacques Martinerie, Bernard Renault, and Francisco J. Varela, 1999, Perception's shadow: Long-distance synchronization of human brain activity", *Nature*, 397: 430–433.

Rorty, Richard, 1970, "Incorrigibility as the mark of the mental", *Journal of Philosophy*, 67: 399–424.

Rosenthal, David M., 1990, "Two concepts of consciousness", *Philosophical Studies*, 49: 329–359

–––, 2001, "Introspection and self-interpretation", *Philosophical Topics*, 28(2): 201–233.

–––, 2005, *Consciousness and Mind*, Oxford: Oxford University Press.

Ryle, Gilbert, 1949, *The concept of mind*, New York: Barnes and Noble.

Sandberg, Kristian, Bahador Bahrami, Ryota Kanai, Gareth Robert Barnes, Morten Overgaard, and Geraint Rees, 2013, "Early visual responses predict conscious face perception within and between subjects during binocular rivalry", *Journal of Cognitive Neuroscience*, 25: 969–985.

Schwitzgebel, Eric, 2002, "A phenomenal, dispositional account of belief", *Noûs*, 36: 249–275.

–––, 2005, "Difference tone training", *Psyche*, 11(6).

–––, 2007, "No unchallengeable epistemic authority, of any sort, regarding our own conscious experience—contra Dennett?", *Phenomenology and the Cognitive Sciences*, 6: 107–113.

–––, 2011a, "Knowing your own beliefs", *Canadian Journal of Philosophy*, 35, supplement 41–62 (*Belief and Agency*, ed. D. Hunter.)

–––, 2011b, *Perplexities of consciousness*, Cambridge, MA: MIT.

–––, 2012, "Introspection, what?", in *Introspection and consciousness*, Declan Smithies and Daniel Stoljar (eds.), Oxford: Oxford.

Scollon, Christie Napa, Ed Diener, Shigehiro Oishi, Robert Biswas-Diener , 2005, "An experience-sampling and cross-cultural investigation of the relation between pleasant and unpleasant affect", *Cognition and Emotion*, 19: 27–52.

Searle, John R., 1983, *Intentionality*, Cambridge: Cambridge.

–––, 1992, *The rediscovery of the mind*, Cambridge, MA: MIT Press.

Shoemaker, Sydney, 1963, *Self-knowledge and self-identity*, Ithaca, NY: Cornell University Press.

–––, 1968, "Self-reference and self-awareness", *Journal of Philosophy*, 65: 555–567.

–––, 1988, "On knowing one's own mind", *Philosophical Perspectives*, 2: 183–209.

–––, 1994a, "Self-knowledge and 'inner sense'. Lecture I: The object perception model, *Philosophy and Phenomenological Research*, 54: 249–269.

–––, 1994b, "Self-knowledge and 'inner sense'. Lecture II: The broad perceptual model", *Philosophy and Phenomenological Research*, 54: 271–290.

–––, 1994c, "Self-knowledge and 'inner sense'. Lecture III: The phenomenal character of experience", *Philosophy and Phenomenological Research*, 54: 291–314.

–––, 1995, "Moore's paradox and self-knowledge", *Philosophical Studies*, 77: 211–228.

Siewert, Charles, 2004, "Is experience transparent?", *Philosophical Studies*, 117: 15–41.

–––, 2012, "On the phenomenology of introspection", in *Introspection and consciousness*, Declan Smithies and Daniel Stoljar (eds.), Oxford: Oxford.

Sirken, Monroe G., Douglas J. Herrmann, Susan Schechter, Norbert Schwarz, Judith N. Tanur, Roger

Tourangeau (eds.), 1999, *Cognition and survey research*, New York: John Wiley and Sons.

Smallwood, Jonathan, and Jonathan W. Schooler, 2006, "The restless mind", *Psychological Bulletin*, 132: 946–958.

Smith, A.D., 2008, "Translucent experiences", *Philosophical Studies*, 140:197–212.

Spener, Maja, forthcoming, "Disagreement about cognitive phenomenology", in *Cognitive phenomenology*, Tim Bayne and Michelle Montague (eds.), Oxford: Oxford University Press.

Stoljar, Daniel, 2004, "The argument from diaphanousness", in *New essays in the philosophy of language and mind*, Maite Ezcurdia, Robert J. Stainton, and Christopher Viger (eds.), Calgary: University of Calgary, 341–390.

Stone, Jeff, and Joel Cooper, 2001, "A self-standards model of cognitive dissonance", *Journal of Experimental Social Psychology*, 37: 228–243.

Summerfield, Christopher, Anthony Ian Jack, and Adrian Philip Burgess, 2002, "Induced gamma activity is associated with conscious awareness of pattern masked nouns", *International Journal of Psychophysiology*, 44: 93–100.

Taylor, Shelley E., and Jonathon D. Brown, 1988, "Illusion and well-being: A social psychological perspective on mental health", *Psychological Bulletin*, 103: 193–210.

Thomas, Nigel, 1999, "Are theories of imagery theories of imagination?", *Cognitive Science*, 23: 207–245.

Titchener, E.B., 1901–1905, *Experimental psychology*, New York: Macmillan.

–––, 1908/1973, *Lectures on the elementary psychology of feeling and attention*, New York: Arno.

–––, 1912a, "Prolegomena to a study of introspection", *American Journal of Psychology*, 23: 427–448.

–––, 1912b, "The schema of introspection", *American Journal of Psychology*, 23: 485–508.

Tong, Frank, Ming Meng, and Randolf Blake, 2006, "Neural bases of binocular rivalry", *Trends in Cognitive Sciences*, 10: 502–511.

Tong, Frank, Ken Nakayama, J. Thomas Vaughan, and Nancy Kanwisher, 1998, "Binocular rivalry and visual awareness in human extrastriate cortex", *Neuron*, 21: 753–759.

Tononi, Giulio, and Christof Koch, 2008, "The neural correlates of consciousness: An update", *Annals of the New York Academy of Sciences: The Year in Cognitive Neuroscience 2008*, 1124: 239–261.

Tononi, Giulio, Ramesh Srinivasan, D. Patrick Russell, and Gerald M. Edelman, 1998, "Investigating neural correlates of conscious perception by frequency-tagged neuromagnetic responses", *Proceedings of the National Academy of Sciences*, 95: 3198–3203.

Tye, Michael, 1995, *Ten problems about consciousness*, Cambridge, MA: MIT.

–––, 2000, *Consciousness, color, and content*, Cambridge, MA: MIT.

–––, 2002, "Representationalism and the transparency of experience" *Noûs*, 36: 137–151.

–––, 2009, *Consciousness revisited*, Cambridge, MA: MIT.

Van Gulick, Robert, 1993, "Understanding the phenomenal mind: Are we all just armadillos?", in *Consciousness: Psychological and philosophical essays*, Martin Davies and Glyn W. Humphreys (eds.), Oxford: Blackwell, 134–154.

Vanman, Eric J., Brenda Y. Paul, Tiffany A. Ito, and Norman Miller, 1997, "The modern face of prejudice and structural features that moderate the effect of cooperation on affect" *Journal of Personality and Social Psychology*, 73: 941–959.

Varela, Francisco J., 1996, "Neurophenomenology: A methodological remedy for the hard problem", *Journal of Consciousness Studies*, 3(4): 330–49.

Vazire, Simine, 2010, “Who knows what about a person? The Self-Other Knowledge Asymmetry (SOKA) model”, *Journal of Personality and Social Psychology*, 98: 281–300.

Velleman, J. David, 2000, *The possibility of practical reason*, Oxford: Oxford University Press.

Watson, John B., 1913, “Psychology as the behaviorist views it”, *Psychological Review*, 20: 158–177.

Wegner, Daniel M., 2002, *The illusion of conscious will*, Cambridge, MA: MIT.

Wegner, Daniel M. and Thalia Wheatley, 1999, “Apparent mental causation”, *American Psychologist*, 54: 480–492.

White, Peter A., 1988, “Knowing more about what we can tell: ‘Introspective access’ and causal report accuracy ten years later”, *British Journal of Psychology*, 79: 13–45.

Williams, Amanda C. de C., Huw Talfryn Oakley Davies, and Yasmin Chadury, 2000, “Simple pain rating scales hide complex idiosyncratic meanings”, *Pain*, 85: 457–463.

Wilson, Timothy D., 2002, *Strangers to ourselves*, Cambridge, MA: Harvard.

Wilson, Timothy D., Samuel Lindsey, and Tonya T. Schooler, 2000, “A model of dual attitudes”, *Psychological Review*, 107: 101–126.

Wittenbrink, Bernd, Charles M. Judd, and Bernadette Park, 1997, “Evidence for racial prejudice at the implicit level and its relationship with questionnaire measures”, *Journal of Personality and Social Psychology*, 72: 262–274.

Wittenbrink, Bernd, and Norbert Schwarz (eds.), 2007, *Implicit measures of attitudes*, New York: Guilford.

Wittgenstein, Ludwig, 1953/1968, *Philosophical investigations*, 3rd edition, G.E.M. Anscombe (translator), New York: Macmillan.

Wollheim, Richard, 1981, *Sigmund Freud*, New York: Cambridge.

–––, 2003, “On the Freudian unconscious”, *Proceedings and Addresses of the American Philosophical Association*, 77(2): 23–35.

Wright, Crispin, 1989, “Wittgenstein's later philosophy of mind: Sensation, privacy, and intention”, *Journal of Philosophy*, 86: 622–634.

–––, 1998, “Self-knowledge: The Wittgensteinian legacy”, in *Knowing our own minds*, Crispin Wright, Barry C. Smith, and Cynthia Macdonald (eds.), Oxford: Oxford.

Wundt, Wilhelm, 1874/1908, *Grundzüge der physiologischen Psychologie* (6th ed.), Leipzig: Wilhelm Engelmann.

–––, 1888, Selbstbeobachtung und innere Wahrnehmung, *Philosophische Studien*, 4: 292–309.

–––, 1896/1902, *Outlines of psychology* (4th ed.), 2nd English ed., Charles Hubbard Judd (trans.), Leipzig: Wilhelm Engelmann.

–––, 1907, “Über Ausfrageexperiments und über die Methoden zur Psychologie des Denkens”, *Psychologische Studien*, 3: 301–360.

Zimmerman, Aaron, 2007, “The nature of belief”, *Journal of Consciousness Studies*, 14(11): 61–82.

Academic Tools

- How to cite this entry.
- Preview the PDF version of this entry at the Friends of the SEP Society.
- Look up this entry topic at the Indiana Philosophy Ontology Project (InPhO).
- Enhanced bibliography for this entry at PhilPapers, with links to its database.

Other Internet Resources

- Implicit Association Test, from Project Implicit, Harvard University.
- Difference Tone Training, Schwitzgebel's (2005) recreation of an introspective training procedure from Titchener's (1901–1905) lab manual.
- Color Wheels; Color Systems, An image of the Munsell color solid can be found at in the pages for the course 2D Design (Art 107), by Curt Heuer at the University of Wisconsin at Green Bay.

Related Entries

behaviorism | belief | Brentano, Franz | consciousness | consciousness: and intentionality | consciousness: higher-order theories | consciousness: representational theories of | consciousness: unity of | delusion | Descartes, René: epistemology | folk psychology: as a theory | folk psychology: as mental simulation | functionalism | Helmholtz, Hermann von | James, William | Kant, Immanuel: view of mind and consciousness of self | mental content: externalism about | mental content: narrow | mental imagery | pain | perception: the problem of | phenomenology | propositional attitude reports | qualia | Ryle, Gilbert | self-consciousness: phenomenological approaches to | self-deception | self-knowledge | Wundt, Wilhelm Maximilian

Stanford | Center for the Study of
Language and Information

Library of Congress Catalog Data: ISSN 1095-5054