

## **Introspection, transparency, and desire**

Michael Roche ([rochmich@isu.edu](mailto:rochmich@isu.edu)), Department of English and Philosophy, Idaho State University, USA

<https://www.ingentaconnect.com/contentone/imp/jcs/2023/00000030/f0020003/art00006>

**ABSTRACT:** The transparency approach to introspection has received much attention over the last few decades. It is inspired by some well-known remarks from Gareth Evans (1982). Although this approach can seem quite plausible as applied to belief (and perhaps perception), philosophers tend to be skeptical that it can succeed for other mental kinds. This paper focuses on *desire*. It lays out in detail a transparency theory of desire introspection and addresses various concerns and objections to such a theory. The paper takes as its launching point Alex Byrne's (2018) influential work on transparency.

**KEYWORDS:** introspection; transparency; desire; desirability; Alex Byrne

### **1 Introduction**

This paper concerns introspection. I take introspection to be a way of acquiring knowledge of one's current mental states without inference from behavioral, circumstantial, and/or psychological evidence about oneself. This characterization is intended to be weak enough to be consistent with extant theories of introspection but strong enough to distinguish introspective knowledge from non-introspective knowledge of various kinds, including knowledge of one's mental states (current or past) inferred from evidence about oneself.

The existence of introspection is controversial. Those who endorse a first-/third-personal symmetry, on which all knowledge of one's mental states comes from mindreading, reject it.<sup>1</sup> Introspection is even more controversial when the above characterization is strengthened. It is often claimed to be both *epistemically privileged* as compared to non-introspective methods and *distinctively first-personal*, that is, suited only for producing knowledge of one's own mental states.<sup>2</sup> Each of these (alleged) features can be spelled out in different ways.<sup>3</sup>

This paper's focus is *desire introspection*. Can we introspect desires? Or are we only able to infer our desires from evidence about ourselves? If we can introspect desires, how does such introspection work? These are difficult questions for which there is little philosophical agreement. To lay my cards on the table, I believe that there *is* desire introspection. I also believe that some (and perhaps all) desire introspection is *transparent* in that it is subserved

---

<sup>1</sup> I take Carruthers (2010, 2011), Cassam (2014), Gopnik (1993), Ryle (1949), and, to a lesser extent, Schwitzgebel (2008, 2012) to hold such a view. Arguably, none goes so far as to maintain a perfect symmetry. Carruthers (2011), e.g., claims that the mindreading system has non-interpretive access to sensory states, including sensory imagery.

<sup>2</sup> See Gertler (2021, §1.1 and §1.2) and Schwitzgebel (2019, §1, §2.1, and §4.1).

<sup>3</sup> The former has been spelled out in terms of safety (Byrne, 2018), reliability (Fernandez, 2013), agency (Moran, 2001), directness (Gertler, 2012), entitlement (Burge, 1996), and grounding (Bar-On, 2004). The latter admits of various strengths depending on how the restriction is interpreted. Doyle (2019) takes it to be metaphysical. Armstrong (1968/1993, p 124) and Lycan (1996, p 49) take it to be merely psychological.

by a world-to-mind introspective method.<sup>4</sup> My primary aim here is to develop and defend a transparency theory of desire introspection.

In Section 2, I sketch the transparency approach to introspection. In Section 3, I describe how that approach can be applied to desire introspection. In each case, I make use of Alex Byrne's (2005, 2011a, 2018) influential work on transparent introspection. The remainder of the paper defends and develops in novel ways a transparency theory of desire introspection.

## 2 Transparency and introspection

The transparency approach to introspection is inspired by some oft-cited remarks from Gareth Evans (1982). He 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?' (p 225)

Evans' idea, put roughly, is that I can know whether I believe that *p* by considering whether it is true that *p*; that is, I can consider the potential belief's content.<sup>5</sup> In this way, questions regarding belief are said to be "transparent to" questions regarding what is the case.

---

<sup>4</sup> The "transparency" label attaches to two quite different views, neither of which I am concerned with here. One concerns self-intimation (Gertler, 2021, §1.1) and the other representational content (Lycan, 2019, especially §3). See also Paul (2014).

<sup>5</sup> One way in which this approximation is rough is that it is silent regarding the possibility that I am unsure whether *p*. See Byrne (2018, p 118) for relevant discussion.

This world-directedness is essential to the transparency approach; transparent introspective knowledge is acquired by attending to non-mental matters. While many philosophers have defended versions of this idea, it has been developed in most detail by Byrne (2005, 2011a, 2011b, 2011c, 2012, 2018).<sup>6</sup>

Because the transparency approach applies most straightforwardly to *belief* introspection, I shall focus on it for purposes of illustration. Evans' comments suggest a method whereby one moves from the judgment that *p* to the judgment that one believes that *p*.<sup>7</sup> Proponents of transparency tend to point to the reliability or safety of this transition; I cannot easily go wrong by self-ascribing the belief that *p* in response to my judgment that *p*. Some even suggest that error is (practically) impossible here.<sup>8</sup>

Consequently, the transparency method is alleged to produce especially secure knowledge. It is also alleged to be distinctively first-personal: in general, ascribing beliefs to *others* in response to one's *own* judgments about the world is risky; many such judgments are highly idiosyncratic (e.g., my coffee is cold).<sup>9</sup> Moreover, when I self-ascribe a belief about

---

<sup>6</sup> Others who have defended transparent introspection, understood broadly, include: Dretske (1994, 1999); Fernández (2013); Gallois (1996); Gordon (2007); and Moran (2001).

<sup>7</sup> I prefer “judgment” over “belief”, for the former more strongly suggests mental *action*.

<sup>8</sup> Byrne (2011b)—following Brueckner (1998)—claims that “... inference from a premiss entails belief in that premiss” (p 206). Each cites Gallois (1996). Error is only “practically” impossible since the belief *could* be lost before the inference is completed. See Byrne (2018, p 104). I take no stand here on whether judgment entails belief or is a kind of belief.

<sup>9</sup> Could I use the method to reliably ascribe beliefs to an *omniscient* being? Only to the extent that my worldly judgments are reliably true. This qualification does not apply when the method is used first-personally, however. I return to this point below.

there being a third world war in response to a judgment about there being a third world war, I do not infer this belief from evidence about myself.<sup>10</sup> For these reasons, the transparency method seems to qualify as introspective, at least as characterized in Section 1.

Relevant here is a point worth emphasizing. Transparent introspective knowledge is claimed to be the product of thinking about the world, *not* thinking about how one takes (or regards) the world to be. How one takes the world to be is a *psychological* matter. The process of self-ascribing a belief on this basis would not be world-to-mind, and so would not be transparent. Nor would it be introspective (as characterized in Section 1), for the belief would be inferred from psychological evidence.

In fact, whether the transparency method is genuinely inferential is unclear. The method differs from ordinary inference in at least two ways. First, its reliability does not depend on the accuracy of the worldly judgments (Byrne 2018, pp 106-108). If I believe that the Earth is flat, I can use the transparency method to truly and non-accidentally self-ascribe this belief. To the extent that the method is highly reliable, this is in virtue of the connection between judgment—whether true or false—and belief.

Second, because the state of the world is largely independent of the state of one's mind—questions about the world and mind concern “different topics” (O'Brien, 2007)—the worldly judgment's content will often (though not always<sup>11</sup>) be very weak evidence for the self-

---

<sup>10</sup> This will not always be the case. If I self-ascribe the belief that I am six feet tall in response to my judgment that I am six feet tall, my self-ascription *is* based on evidence about myself (for I *am* six feet tall). Yet this fact is inessential to the method's reliability.

<sup>11</sup> Suppose that I use the transparency method to self-ascribe the belief that it is raining outside. Because people tend to have accurate beliefs about the current weather in their environment,

ascription (Byrne 2018, pp 105-106). This is so even when the judgment's content is known.<sup>12</sup> When I knowingly judge that my car's gas tank is low, the judgment's content is not strong evidence for my believing that the tank is low. (Often the tank is low without my believing so.) Despite this, I can use the transparency method to truly and non-accidentally self-ascribe the belief that the tank is low in response to that judgment.<sup>13</sup>

Whether the transparency method is genuinely inferential is to some extent a red herring.<sup>14</sup> More important, I think, is that we acknowledge that the method differs in significant ways from ordinary inference, including our inferences about *others'* mental states. In this way, the transparency approach maintains a significant first-/third-personal asymmetry with respect to the attribution of mental states.<sup>15</sup>

---

if it truly is raining outside, then arguably my self-ascription *is* based on adequate evidence. However, as with the previous footnote, this is inessential to the method's reliability.

<sup>12</sup> If one's evidence is one's knowledge (Williamson 2000), then when the worldly judgment is false, one will not reason from evidence, period.

<sup>13</sup> These points about evidence are related to what Byrne (2005, 2012, 2018) calls the "puzzle of transparency". It is forcefully expressed by Dretske (2003, 2012).

<sup>14</sup> Byrne calls the method "inferential" but notes that it is not an inference from "adequate evidence" (2018, p 123). For further discussion, see Valaris (2011) and Setiya (2012).

<sup>15</sup> Certain logically trivial inferences resemble the transparency method. Consider, e.g., the inference from P to (Q or not-Q). First, given that the conclusion is tautologous, this inference cannot easily go wrong (it will always result in truth). Second, this is so regardless of whether the premise is true. Third, the premise is not evidence for the conclusion, at least not in any interesting sense. I thank an anonymous referee for pointing this out. Still, because such inferences are far from ordinary, the above point seems to stand.

An attractive feature of transparent introspection is its *metaphysical modesty*.<sup>16</sup> Because transparent introspective knowledge is acquired by attending to the world, the mechanisms/capacities that make possible first-order thought and perception go a long way in securing our introspective capacities. Also required are mental state concepts and the capacity to make introspective judgments in response to worldly judgments.<sup>17</sup>

This modesty sets transparency theories apart from immodest views of introspection, such as inner-sense and acquaintance theories. Inner-sense theories posit specialized psychological mechanisms tasked with monitoring one's mind. Acquaintance theories posit metaphysically direct acquaintance relations holding between persons and (some of) their mental states. Although many are skeptical of the transparency approach to introspection, the mechanisms/capacities it describes are not in doubt.<sup>18</sup>

A somewhat related issue that deserves more attention than it receives concerns our *use* of transparency methods.<sup>19</sup> Why and how do we develop a disposition to use them? Although I can only speculate here, I am inclined to think that we do so out of need and/or usefulness.

---

<sup>16</sup> Byrne calls this “economy” (2018, p 14). See Byrne (2018, pp 112-116) for a discussion of economy with respect to the transparency method for belief introspection.

<sup>17</sup> The transparency approach is naturally paired with a view on which such concepts are acquired independently of introspection, e.g., Gopnik and Wellman's (1992) “child scientist theory”. The capacity to reason transparently presumably falls out of a more general capacity for reasoning. See Byrne's discussion of “epistemic rules” (2018, pp 100-102).

<sup>18</sup> Inner-sense theorists include Armstrong (1968/1993, 1981), Goldman (2006), Lycan (1996), and Nichols and Stich (2003). Acquaintance theorists include Chalmers (2003) and Gertler (2001, 2012); see Duncan (2021) for a general discussion of acquaintance.

<sup>19</sup> Byrne, to his credit, says a bit about this (2018, pp 113-114).

Specifically, what is needed and/or useful is the ability to know our mental states without inference from evidence about ourselves. We naturally discover transparency methods as a solution. Perhaps aiding us in this discovery is an implicit recognition that worldly matters are relevant to which attitudes we *ought* to have.<sup>20</sup> We then come to think about worldly matters when questions about our attitudes arise.<sup>21</sup>

### 3 Transparency and desire introspection – Byrne’s theory

The transparency approach to belief introspection claims that beliefs can be known by considering what is true. Yet this is not the case for other mental kinds. In particular, *desires* cannot be known in this way. To regard things as being one way while wanting them to be a different way is not at all uncommon. Extending the approach from belief to desire thus requires adjustments.<sup>22</sup>

If questions regarding desires are not transparent to questions regarding truth, what might they be transparent to? A natural candidate is *desirability*. Echoing Evans’ earlier comment, it seems that I can answer the psychological question, ‘Do I want there to be a third world

---

<sup>20</sup> Such normativity plays a role in Moran’s (2001) transparency theory.

<sup>21</sup> Alternatively, perhaps some mental kinds (e.g., belief and perception) are “transparency friendly”, making it practically irresistible to self-ascribe them transparently. This habit might then be naturally extended to other mental kinds.

<sup>22</sup> Some critics of transparency grant its attractiveness as applied to belief (and perhaps perception) while maintaining that it simply cannot work for other mental kinds. Extending the approach is of great importance to transparency theorists with generalist ambitions. Byrne (2018) represents the most thorough attempt to meet this challenge. His critics include Ashwell (2013), Keeling (2020), and Paul (2015, 2020).

war?,' by instead answering the non-psychological question, 'Would a third world war be desirable?' This, at any rate, is the suggestion explored in this paper. It is also the route that Byrne (2018, 2011a) takes. His work will serve as my launching point.

Byrne alleges that there is *alignment* between one's desirability judgments and one's desires. Focusing on options for actions (e.g., going to the Sushi bar or the Indian restaurant), he writes that: "... one's desires tend to line up with one's *beliefs* about the desirability of the options, whether or not those beliefs are actually true ..." (2018, p 162, original emphasis). Any such alignment is of course imperfect; we can judge that an option is desirable without desiring it.<sup>23</sup> Byrne acknowledges this, claiming that perfect alignment is not required.

Because transparency methods are world-to mind, the relevant sense of "desirability" must be world-directed. Accordingly, Byrne has in mind "... the *Oxford English Dictionary* sense of having 'the qualities which cause a thing to be desired: Pleasant, delectable, choice, excellent, goodly'" (2018, p 160). Crucially, these qualities are (potential) qualities of external matters, in particular of the options (e.g., going to the Sushi bar). Moreover, even if these qualities are *dispositional* in that they are "... qualities which cause [an action] to be desired ...", I can judge that an option possesses such qualities without judging that *I* have any particular desires.<sup>24</sup> Desirability judgments, so understood, are world-directed judgments.

---

<sup>23</sup> This marks a difference between the transparency methods for belief and desire. The claim that judging that p entails believing that p is much less controversial than the corresponding claim for desirability judgments and desires. Indeed, some of Byrne's critics have made this very point: e.g., Doyle (2019, p 517) and Paul (2020, p 483). See Peacocke (1999) and Zimmerman (2006) for opposing views regarding belief.

<sup>24</sup> See Byrne (2018, pp 162-164) for discussion of other worries about circularity.

Byrne's transparency method for desire introspection thus takes one from a judgment that an option is desirable (or undesirable) to the judgment that one desires (or does not desire) that option.<sup>25</sup> Byrne takes this method to be highly reliable in virtue of the (alleged) alignment between desirability judgments and desires.<sup>26</sup> Because there is much less alignment between one's desirability judgments and *others'* desires, the method is highly reliable only when used first-personally. Additionally, the method does not involve inference from behavioral, circumstantial, and/or psychological evidence about oneself. Desires are self-ascribed in response to judgments about the desirability of options. For these reasons, the method seems to qualify as introspective according to the characterizations from Section 1.<sup>27</sup>

A final feature of Byrne's account worth noting concerns *defeasibility*. As already noted, the method does not require perfect alignment between desires and desirability judgments.

Byrne describes a case where these come apart:

Lying on the sofa, wallowing in my own misery, I know that going for a bike ride by the river is a desirable option. The sun is shining, the birds are twittering, the exercise and the scenery will cheer me up; these facts are easy for me to know, and my torpor does not prevent me from knowing them. (2018, p 161)

Despite judging that going for a bike ride is desirable, he does not want to go for a bike ride.

Byrne claims that the misalignment here is harmless, for he will simply *bypass* the transparency method in these circumstances. The method will be bypassed, he claims, given

---

<sup>25</sup> The parentheticals denote a variation of the method that Byrne (2018, p 166) endorses.

<sup>26</sup> Byrne speaks of "safety", not "reliability". However, because nothing in this paper turns on this difference, I shall use the latter, somewhat more straightforward, notion. See Sosa (1999) and Williamson (2001) for discussion of safety.

<sup>27</sup> It also differs from ordinary inference in the two ways noted in Section 2.

that he knows that he intends to continue lying on the couch—an option that he takes to be neither desirable nor all-things-considered better than the alternative.<sup>28</sup> I return to the matter of bypassing the transparency method in Section 4.5.

## 4 Transparency and desire introspection - Elaborations and defenses

Although Byrne's discussion is instructive, it is somewhat abbreviated and consequently underdeveloped.<sup>29</sup> Much more can be said on behalf of the transparency approach to desire introspection. I supply additional arguments and address numerous worries in what remains.

### 4.1 The contents of desirability judgments

I begin by considering the *contents* of the relevant desirability judgments. Additional points concerning these contents are made in sections 4.2 and 4.4.

Because Byrne focuses on desires for actions, he takes the corresponding desirability judgments to concern options for action. Yet I think we should broaden the scope so that desires are for (*possible*) *states of affairs*.<sup>30</sup> Although desired states of affairs will sometimes include my performing certain actions (e.g., I want to go to the gym), sometimes they will not (e.g., I want the candidate to win).

More significantly, we should understand desirability judgments to be *all-things-considered* judgments. This is because we rarely judge as desirable all aspects of a possible state of affairs. Is my learning to play the tuba desirable? Yes and no. There are various

---

<sup>28</sup> Byrne also defends a transparency theory of knowing one's intentions (2018, 2011b). See Boyle (2011) and Paul (2015) for criticism of it.

<sup>29</sup> This is not to say that I have explained all aspects of it. I have not.

<sup>30</sup> See Schroeder (2020, §2.1) for further discussion.

respects in which it is: I value music; I enjoy playing instruments; I know a great teacher; etc. And there are various respects in which it is not: I lack the required time; I prefer string over brass instruments; the weight of the tuba would aggravate my back problems; etc. Without interpreting the question as concerning an all-things-considered judgment, it is ill-defined.<sup>31</sup>

As a matter of fact, I judge that my learning to play the tuba is undesirable all things considered, for I judge that it is undesirable in more respects than not and/or assign greater weight to those respects. If I inquired as to whether I desire to learn to play the tuba, the transparency method would have me conclude that I do not. This result seems correct.

By stressing that the relevant world-directed judgments are all-things-considered judgments, I do not mean to suggest that they are products of labored consideration. We are typically able to make such judgments relatively quickly and with ease. *How* we can do this is unclear. But *that* we can do this should be uncontroversial. If I were to ask my partner whether it would be desirable for her to learn to play the tuba, she would quickly and easily answer that it is not. And she will do so despite acknowledging that there are certain respects in which her learning to play the tuba is desirable. Although she does not take the time to explicitly consider all things, her judgment is apparently an all-things-considered judgment.

Finally, while I have followed Byrne in characterizing the relevant judgments in terms of *desirability*, I do not regard this as essential to the approach. Perhaps characterizing the judgments in terms of *goodness* would do just as well. What matters, I think, is that the relevant states of affairs are positively evaluated. This is not to claim that desires *are* positive

---

<sup>31</sup> This is not necessarily a departure from Byrne's theory. His response (2018, p 166) to an objection from Ashwell (2013) suggests that he understands desirability judgments in this way. Yet he is not explicit about this when presenting his theory.

evaluative judgments, however.<sup>32</sup> An identification of this kind is not needed to maintain that self-ascribing desires in response to positive evaluative judgments is highly reliable.

#### 4.2 Desirability judgments and world-directedness

The state of affairs from the previous example involves me; *I* am the one who would be learning to play the tuba. This will often, though not always, be the case. Many desires are egocentric.<sup>33</sup> Importantly, this is compatible with the method's being world-directed in the way that transparency requires.

When I judge that my learning to play the tuba is undesirable all things considered, I am commenting on the state of affairs in question. Specifically, I am judging that it—my learning to play the tuba—lacks certain qualities: it is not “[p]leasant, delectable, choice, excellent, goodly”.<sup>34</sup> I am *not* judging that I regard that state of affairs to be undesirable, for that judgment would be about my mind. Although we can make judgments of the latter kind, my point is simply that in so doing, we are not using the transparency method.

Having said this, we should admit that a person's overall state of mind will factor into her desirability judgments. Judgments of any kind are products of minds, after all. I judge that my taking tuba lessons is undesirable. Yet if my overall mental state were different, then I might judge differently. If, for example, I lacked the aforementioned preference for string

---

<sup>32</sup> Nor is it to claim that desires *are* positive evaluative perception-like seemings, as defended, e.g., by Stampe (1987) and Oddie (2005).

<sup>33</sup> Stampe (1987) discusses a related matter (pp 374-376).

<sup>34</sup> Recall that even if these qualities are *dispositional* in that they are defined in terms of their effects on minds (e.g., causing desires), a person can judge that such a quality is present without judging that she is so affected.

over brass instruments, then I might assign different weights to the considerations for and against the desirability of my taking tuba lessons. This fact is consistent with the judgment's being world-directed. That a person's judgments are a function of his overall mental state does not, by itself, make the contents of those judgments mentalistic.

There is, however, a potentially more troubling case. Rather than being merely causally impactful, a mental state might be explicitly cited as a consideration in one's all-things-considered desirability judgment. Suppose that I explicitly cite my preference for string instruments as a consideration against the desirability of my taking tuba lessons. This suggests that my belief that I have this preference—a belief about my mind—is doing some work. If so, then my judgment is *not* purely world-directed, after all. That is the worry.

My response to this is fourfold. First, desirability judgments do not *require* thinking about one's mental states. Because desirability judgments—properly understood—concern the qualities of possible states of affairs, strictly speaking, one's attention can focus solely on external matters. Whether one thinks about one's mind when so judging is optional. Second, thinking about one's mind when making a judgment does not make that judgment's content mentalistic. Suppose that I judge that a bridge is unsafe due, in part, to my belief that I feel fear when crossing it. Still, my judgment that the bridge is unsafe is about *the bridge*.

Third, so long as such considerations do not involve beliefs about one's desires, there is no obvious problem of circularity. The relevant belief in the tuba case concerns a *preference*, not a desire. Moreover, that preference is not directly about my taking tuba lessons. If and when I judge that I do not desire to take tuba lessons, I thus make a *new* judgment. Fourth, any beliefs about one's mental states that figure in one's considerations of desirability might

themselves be products of transparency methods.<sup>35</sup> When they are, the resulting picture is one that transparency theorists should find satisfying, not troubling.<sup>36</sup>

#### 4.3 The transparency method and awareness

The transparency method produces knowledge of our desires. It does not give us knowledge (or even belief) of our desirability judgments. Nor does it give us knowledge (or even belief) of the causal connections between our desirability judgments and desire self-ascriptions. The transparency approach to desire introspection is thus compatible with our being unaware of self-ascribing desires in response to desirability judgments. Byrne is clear on these points (2018, p 102, p 114), admitting that our use of the transparency method

“... is *not* evident ...” (p 114). Far from being a problem, however, he notes that introspective knowledge is often described as appearing *groundless* (2018, p 123).

That much of our reasoning occurs unconsciously, without our awareness, is not controversial.<sup>37</sup> Although the transparency method differs from ordinary inference in ways already described, it is at least inference-like. For these reasons, it is certainly legitimate to claim that we use the transparency method without awareness.

---

<sup>35</sup> Byrne argues that introspection is *uniform* (2018, pp 157-158). Consequently, for him, if preferences are introspectable, they are introspected transparently. Such unification is not built into the transparency approach, however. Non-transparent alternatives—whether introspective or non-introspective—will be discussed in Section 4.5. For criticism of Byrne’s argument for introspective uniformity, see Roche (forthcoming).

<sup>36</sup> Indeed, even beliefs about one’s *desires* could play a legitimate role when considering desirability, if those beliefs were themselves products of transparent introspection.

<sup>37</sup> See Evans and Frankish (2009) for discussion of unconscious inference and processing.

Still, it would be nice to have some explanation of *why* this is so. Part of the reason, I suggest, is that we sometimes self-ascribe desires in virtue of *prior* desirability judgments. When this occurs, a new judgment is not made, nor is desirability considered anew. Although the prior judgment is doing work—it is in virtue of this judgment that the desire is self-ascribed—its role will be masked. Consequently, the self-ascription may appear groundless.

This certainly seems to occur for belief. Some matters are, to my mind, settled. I have often considered whether there will be a third world war. If asked whether there will be a third world war, I might simply make use of a prior judgment on the matter (not considering the matter anew). If I wondered whether I *believe* that there will be a third world war, I might very quickly and with little conscious effort self-ascribe a belief in virtue of a prior judgment. If so, I will have self-ascribed the belief transparently. Yet no new judgment was made. Nor did I consider the prospect of war.

Admittedly, this suggestion is limited. In some cases, desirability (or truth) will be considered anew. Perhaps I have never considered the matter in question, or perhaps I have acquired new information that I take to be relevant to that matter. Still, and as already noted, such judgments can be made quickly and with little conscious deliberation. Even here, then, being unaware of having self-ascribed a desire transparently would not be surprising.<sup>38</sup>

---

<sup>38</sup> When self-ascribing desires in response to prior desirability judgments, there is a risk that those judgments will have fallen out of step with one's desires. This suggests that the transparency method is most reliable when desirability judgments are "fresh". But this is compatible with the method's also being reliable when prior desirability judgments are used. Whether it is reliable in such circumstances depends on the stability of desires, as well as the conditions under which prior desirability judgments are used.

#### 4.4 Alignment between desires and desirability judgments

Transparent desire introspection will be reliable only if there is a sufficient degree of alignment between desires and all-things-considered desirability judgments. Making a positive case for such alignment is difficult, however. Indeed, Byrne does little more than assert that such alignment obtains. Things would be different if there were a *constitutive* connection between desires and desirability judgments. Yet I am unwilling to embrace such cognitivism (neither, it seems, is Byrne). Still, much can be said in favor of alignment.

First, we must keep in mind that *perfect* alignment is not required. The transparency method need not be infallible to produce introspective knowledge. We can grant that a person will sometimes judge that a state of affairs is desirable (all things considered<sup>39</sup>) without desiring it. And vice versa.<sup>40</sup>

Second, cases of misalignment are quite intriguing and likely overshadow more commonplace cases of alignment. In fact, their intrigue might be due to their relative infrequency. To combat this, consider some perfectly ordinary examples of alignment: (i) Alonna desires, and judges as desirable, earning her master's degree in business; (ii) Ben desires, and judges as desirable, his children having good health; (iii) Chandra desires, and judges as desirable, having a competent car mechanic; (iv) Donovan does not desire, nor judge as desirable, getting laid off from his job; (v) Elizabeth does not desire, nor judge as desirable, having mayonnaise on her sandwich; and (vi) Francis does not desire, nor judge as desirable, the end of sexism in hiring practices. (More on this last case below.)

---

<sup>39</sup> For the most part, I shall drop this qualification in what follows.

<sup>40</sup> Judging that p is desirable without desiring that p is not technically an instance of akrasia. Akrasia involves *acting* contrary to what one judges to be best (Stroud and Svirsky, 2021).

The fact that cases like (i)-(vi) are so unremarkable suggests that a given person likely has many desires that align with his or her (potential) desirability judgments, and likely lacks many desires that would not align with such judgments. A natural explanation for this is that a person's desirability judgments are *influenced by* many of the same psychological factors that influence her desire attitudes; that they are influenced by a common set of factors increases the probability that they align. Such a connection would be *causal*, not constitutive.

Consider my desire to watch *Severance* tonight. I enjoy the show's relevance to, and illustration of, philosophical issues about personal identity. Plausibly, my interest in such issues is part of the reason that I desire to watch the show. It is also plausible that this interest is part of the reason that I judge as desirable my watching the show tonight; the show's connection to philosophy is one of the respects in which I judge the state of affairs to be desirable. That there is alignment between my desire and my desirability judgment is not coincidental. Neither, though, is there a constitutive link between the two.

Third, we must be sensitive to the specificity of desirability judgments. Suppose that I judge that eating lots of fatty food, *generally*, is undesirable. This need not prevent me from using the transparency method to know that I desire to eat lots of fatty food *this evening*. We can easily imagine circumstances where I would judge *this* state of affairs to be desirable: suppose that a restaurant is bringing back, for one night only, my favorite yet very fatty entrée. If I were using the transparency method, I would self-ascribe a desire to eat lots of fatty food tonight, but I would not self-ascribe a desire to eat lots of fatty food generally.

This is related to the earlier discussion of egocentricity (Section 4.2). As I argued there, egocentricity is compatible with the world-directedness of the transparency approach. When I judge as desirable (i.e., “[p]leasant, delectable, choice, excellent, goodly”) my eating lots of fatty food tonight, I am not commenting on my psychology. The judgment's content concerns

the world. And this is so regardless of whether my judgment is *influenced* by my beliefs, desires, and values in ways previously illustrated.

Fourth, desirability judgments, like judgments of any kind, are fallible. They can thus align with desires for the undesirable.<sup>41</sup> Return to Francis. I trust that most readers will regard his judgment to be mistaken; the end of sexism in hiring practices *is* desirable. Because Francis is disposed to err in this kind of way, his judgment will align with his desire for the continuation of sexism in hiring practices. That we sometimes desire the undesirable is thus compatible with such desires being known transparently.

Fifth, and finally, there are reasons to distinguish desires from urges. Arguably, urges, but not desires, are necessarily conscious and occurrent. And, arguably, desires, but not urges, are at least somewhat responsive to reasons. Jordi Fernández (2013, p 85) offers an additional difference: one can have an urge for something that one cannot, due to a conceptual limitation, desire; he gives the example of a child who has a sexual urge despite lacking the concept of sexual intercourse. The relevance of this distinction is easily demonstrated.

Sinclair appears to want to have a cigarette during his upcoming work break: he repeatedly and anxiously checks his watch; he often peers over at his co-worker's purse that he knows contains cigarettes; he slips and says "I'd love to have a cigarette" when meaning to say "I'd love to have a coffee"; the visual image of a cigarette dominates his mind; he tokens in inner speech the sentence "I need a cigarette"; etc. Yet Sinclair does not judge that his having a cigarette is desirable all things considered, for he has recently decided that he urgently needs to quit. So described, this is an apparent case of misalignment. Another

---

<sup>41</sup> I do not mean to assume realism about desirability or value here. The point is meant to be compatible with non-realist views of desirability/value.

possibility suggests itself, however. Perhaps Sinclair lacks the desire in question but has a strong urge for nicotine. (I return to this case shortly.)

Admittedly, I have not provided a definitive case for alignment. Those doubtful that there is enough alignment to sustain transparent desire introspection have not been conclusively defeated. Collectively, though, I hope that the above points have weakened such doubts.

#### 4.5 Alternative methods – Bypassing/overriding the transparency method

Some will no doubt be suspicious of the distinction just drawn between urges and desires. We often treat them as equivalent in speech, after all. Sinclair might say, “despite my health issues and better judgment, I really *want* a cigarette!” I only meant to illustrate one way in which some apparent misalignment might be explained away. If that way fails—either in a particular case or across the board—so be it. The transparency approach can tolerate some misalignment.

Still, something remains to be explained. As just suggested, Sinclair might self-ascribe a desire for a cigarette while sincerely judging that his having one is undesirable all things considered. Regardless of whether that self-ascription is true, it conflicts with his desirability judgment. Surely, Sinclair is not alone here. We sometimes self-ascribe desires in ways that conflict with the transparency method. *This* is what needs to be explained. Why is the transparency method not blindly followed whenever we self-ascribe desires?

The answer to this question, I think, lies in recognizing that humans have numerous resources to draw upon when thinking about their minds, introspection being just one. Transparency theorists should acknowledge alternative, non-introspective ways of self-ascribing desires. We can, and do, apply to ourselves our ways of ascribing desires to

others.<sup>42</sup> These ways are inferential, relying on evidence about behavior, circumstances, and psychology; when used on ourselves, we are thus engaged in inference, not introspection.<sup>43</sup>

My suggestion, in short, is that departures from the transparency method can be explained by appeal to such alternatives. The transparency method will sometimes be overridden by, or bypassed in favor of, these alternatives. Before explaining how this might work, it will be helpful to consider such alternatives in more detail.

#### *4.5.1 Alternatives to transparency*

The most obvious non-introspective way of self-ascribing desires is by crude Rylean inference (Ryle, 1949). This is simple inference from evidence about one's behavior and circumstances. Presumably, such inference is aided by folk-psychological background information or theory. A more sophisticated brand of Ryleanism allows for psychological evidence about oneself. Such evidence might itself be acquired by inference from behavioral and circumstantial evidence, in which case the inference remains straightforwardly Rylean.

---

<sup>42</sup> Goldman (2006) defends an inner sense theory of introspection while admitting that we also mindread ourselves. This admission is needed to accommodate confabulation. He calls this a "dual method theory" (p 232). For discussion of confabulation, see Carruthers (2010).

<sup>43</sup> Because introspection is not suited for knowing *others'* minds, it should not be surprising that we have non-introspective ways of ascribing mental states. Why do we *also* have introspection (if we do)? There are numerous possibilities. If introspection is metaphysically modest, as described in Section 2, then there is no mystery: it is simply a byproduct of our more general capacities for thinking about and perceiving the world. Alternatively, there could have been (direct) natural selection for introspection given certain advantages that it has over non-introspective ways of self-ascribing mental states.

There is, however, a kind of inference from psychological evidence that is less obviously Rylean. This is inference from *phenomenological evidence*. Such evidence concerns the phenomenal (or felt) qualities of mental states/events. Suppose, for example, that Sinclair's consciousness is dominated by cigarette-related visual imagery, that he experiences anxiety upon imagining not having a cigarette, and that he repeatedly utters in inner speech, "I want a cigarette", etc. Based on his awareness of the phenomenal qualities of these states/events, he might then self-ascribe a desire for a cigarette.<sup>44</sup>

Notice that the phenomenology just described belongs to Sinclair's visual imagery, anxiety, and inner speech. It does not belong to his desire. Sinclair's self-ascription is thus inferential, not introspective. It is inferred from evidence. Whether that evidence is itself introspected or inferred is irrelevant. Either way, he does not introspect his *desire*.

Might desires themselves have phenomenal qualities? This is unclear. Arguably, only occurrent states/events have such qualities.<sup>45</sup> But suppose that desires are sometimes occurrent and possessing phenomenal qualities. Self-ascribing desires on the basis of such qualities would seem to be a kind of non-transparent desire introspection.

---

<sup>44</sup> Lawlor (2009) defends a *causal self-interpretation* account according to which we infer desires as the causes of various internal promptings (e.g., sensations, sentences in natural language, and sensory imagery). This is quite similar to the kind of inference just described. See also Carruthers (2011) and Cassam (2014). Interestingly, as Byrne (2018, p 198) points out, Ryle (1949, p 28, p 173) seems to allow that awareness of our inner speech can aid us in inferring our mental states.

<sup>45</sup> On the occurrent/non-occurrent distinction as applied to desires, see Schroeder (2020, §2.4). Some philosophers are skeptical of occurrent attitudes in general (e.g., Carruthers, 2010), and of occurrent desires in particular (e.g., Hulse, Read, and Schroeder, 2004).

I wish to remain neutral as to whether desires have phenomenal qualities, and thus as to whether the kind of desire introspection just described is possible.<sup>46</sup> Clearly, though, desires not possessing phenomenal qualities cannot be introspected in this way. Assuming that desires of this kind can be introspected, a different introspective method—one not based in phenomenology—is needed. Desire introspection by the transparency method fits the bill.<sup>47</sup>

#### *4.5.2 Bypassing/overriding the transparency method*

Suppose that Sinclair does not blindly follow the transparency method. He instead self-ascribes a desire for a cigarette. I suggested above that in cases like this, the transparency method is overridden by, or bypassed in favor of, an alternative method for self-ascribing desires. But what exactly does this mean?

Consider, first, bypassing the method. In general, this can occur when one has evidence that one has a particular desire. Earlier, I described potential behavioral, circumstantial, and psychological evidence available to Sinclair suggesting that he *wants* a cigarette. Given this, the introspective question “Do I want a cigarette?” might not even arise for him. Or if it does,

---

<sup>46</sup> Strictly speaking, the transparency theory defended in this paper is compatible with there being additional methods of desire introspection, whether they be transparent or non-transparent. Although I am inclined to think that all desire introspection is transparent, I shall not insist on this point here.

<sup>47</sup> These points are relevant to acquaintance theories of introspection, which I briefly described at the end of Section 2. Presumably, one can be acquainted with a mental state/event only if it possesses phenomenal qualities.

it might be quickly answered based on that evidence. In either case, he will not engage the transparency method.<sup>48</sup>

Rather than being bypassed, the method might instead be overridden. Imagine that Sinclair uses the transparency method and, as a result, comes to believe that he does *not* want a cigarette. Given the available evidence pointing in the opposite direction, however, he infers that he *does* want a cigarette. He ultimately resolves the contradiction by giving up the former belief. He resolves the contradiction this way (rather than the other) because he gives greater credence to the latter belief. Overriding the transparency method at an earlier stage is also possible. Perhaps Sinclair's evidence-based belief that he wants a cigarette prevents him from following through with the transparency method. Although he engages the transparency method, that belief prevents him from concluding that he does not want a cigarette.

That an introspective method can be overridden by, or bypassed in favor of, an inferential/non-introspective method should not be surprising or troubling. Although introspective methods are, by definition, epistemically superior to non-introspective methods, they are fallible. Moreover, the products of introspection do not come labeled as such; an individual who sees fit to ignore or override the results of introspection need not take herself to be doing so. When thinking about one's mind, one will sometimes be pulled in competing and conflicting directions. To avoid contradiction, choices must be made.

## **5 Concluding remarks - Transparency and misalignment**

Despite all that I have said in favor of transparent desire introspection, I suspect that some readers will be unable to see past cases of misalignment. These are of two types. One can

---

<sup>48</sup> Interestingly, if the case is altered by removing such evidence, it becomes more natural to suppose that he will self-ascribe a desire that aligns with his desirability judgment.

desire that which one takes to be undesirable. And one can fail to desire that which one takes to be desirable. I shall close with cases of each, demonstrating how the theory defended here has the resources to handle them.

Lauren Ashwell (2013) discusses a case of the second type. It is actually a variant of Byrne's cycling case (described in Section 3). We are to imagine that the protagonist judges that *both* of his options—cycling and lounging on the couch—are desirable. Regarding the latter, we are to suppose that he judges that "... the couch is comfortable, and staying there takes less effort than getting up" (p 253). Because neither option is judged to be undesirable, Byrne's explanation for why the transparency method is bypassed does not apply. Still, Ashwell suggests, the protagonist might know that he does not desire to go cycling.<sup>49</sup>

At least two responses are available to transparency theorists. Consider, first, the discussion from Section 4.1. Although Ashwell describes the couch dweller as judging that his going cycling is desirable, he might instead merely judge this to be desirable in certain respects, ultimately judging that his going cycling is undesirable all things considered. His apathetic condition might causally contribute to this judgment. So described, the case poses no trouble for the theory.<sup>50</sup>

---

<sup>49</sup> This knowledge, Ashwell suggests, is due to that option's *appearing* undesirable to the protagonist (pp 254-255). Importantly, such appearances are not judgments. Partly for this reason, transparency theorists will be reluctant to appeal to them.

<sup>50</sup> Byrne's (2018) reply to Ashwell is somewhat similar to my first response. He writes that: "[c]ycling can strike one as having significant positive features, yet they can all be trumped by a negatively appearing feature—that cycling will be tiring, say" (p 166). However, this reply does not address the possibility that the protagonist judges that cycling *is* desirable all things considered. My second response addresses this possibility.

Another response draws on the discussion from Section 4.5. Compatible with Ashwell's description of the case is the possibility that there is behavioral, circumstantial, and/or psychological evidence suggesting that the protagonist does *not* desire to go cycling. If so, then his knowledge that he does not desire to go cycling—along with his unwillingness to self-ascribe a desire for cycling—might be explained by inference from such evidence.

Gary Watson (1975, p 210) provides a well-known case of the first type of misalignment. A frustrated mother (briefly) wants to drown her screaming baby in the bath. Despite *not* judging that drowning her baby is desirable, the mother wants to do so and knows this. This case is similar to—though more emotionally charged than—the case of Sinclair. The points made about that case thus apply here as well.

First, it might be that the mother does not want (and so does not know that she wants) to drown her baby. Rather, she merely has an urge to do so.<sup>51</sup> Second, even if she does want to drown her baby, her knowledge of this desire might be explained by an alternative method. The desire in question would likely be accompanied by a range of phenomenology: intense emotions, sensations, and various kinds of imagery. Accordingly, her knowledge of this desire might be inferred from phenomenological evidence.

I should stress that there is no shame in acknowledging the exercise of non-introspective methods for self-ascribing mental states. That we can, and do, use such methods on ourselves is not in doubt. Yet there remains the conviction, at least for many, that we also have a more direct way of knowing our mental states. This way, of course, is introspection. I have argued that transparent introspection—attractive to many due to its metaphysical modesty—can be

---

<sup>51</sup> Watson entertains this possibility but suggests that desiring without valuing (or judging as desirable) is possible, even if this is not such a case.

fruitfully applied to desires. When fully spelled out and defended, transparent desire introspection is quite promising.

\*Many people have provided valuable feedback on previous versions of this paper. Special thanks go to Brie Gertler, Sarah Paul, and Danielle Wylie. Josh May and Shannon Spaulding provided helpful comments on an early version of this paper presented at the Southern Society for Philosophy and Psychology. Thanks as well to the referees at this journal.

## References

- Armstrong, D.M. (1968/1993) *A Materialist Theory of Mind*, New York: Humanities Press.
- Armstrong, D.M. (1981) *The Nature of Mind and Other Essays*, Ithaca, NY: Cornell University Press.
- Ashwell, L. (2013) Deep, dark...or transparent? Knowing our desires, *Philosophical Studies*, **165**, pp. 245-256. [10.1007/s11098-012-9950-3](https://doi.org/10.1007/s11098-012-9950-3)
- Bar-On, D. (2004) *Speaking My Mind: Expression and Self-Knowledge*, Oxford: Oxford University Press.
- Boyle, M. (2011) Transparent self-knowledge. *Proceedings of the Aristotelian Society Supplementary*, **85**, pp. 223-241. [10.1111/j.1467-8349.2011.00204.x](https://doi.org/10.1111/j.1467-8349.2011.00204.x)
- Brueckner, A. (1998) Moore inferences, *Philosophical Quarterly* **48**, pp. 366–369. [10.1111/j.0031-8094.1998.00107.x](https://doi.org/10.1111/j.0031-8094.1998.00107.x)
- Burge, T. (1996) Our entitlement to self-knowledge, *Proceedings of the Aristotelian Society*, **96**, pp. 91–116. [10.1093/aristotelian/96.1.117](https://doi.org/10.1093/aristotelian/96.1.117)
- Byrne, A. (2005) Introspection, *Philosophical Topics* **33** (1), pp. 79-104. [10.5840/philtopics20053312](https://doi.org/10.5840/philtopics20053312)

- Byrne, A. (2011a) Knowing what I want, in Liu, J. & Perry, J. (eds.) *Consciousness and the Self: New Essays*, Cambridge: Cambridge University Press.
- Byrne, A. (2011b) Transparency, belief, intention, *Proceedings of the Aristotelian Society Supplementary*, **85**, pp. 201-221. [10.1111/j.1467-8349.2011.00203.x](https://doi.org/10.1111/j.1467-8349.2011.00203.x)
- Byrne, A. (2011c) Knowing that I am thinking, in Hatzimoysis, A. (ed.) *Self-Knowledge*, Oxford: Oxford University Press.
- Byrne, A. (2012) Knowing what I see, in Smithies, D. & Stoljar, D. (eds.) *Introspection and Consciousness*, Oxford: Oxford University Press.
- Byrne, A. (2018) *Transparency and Self-Knowledge*, New York: Oxford University Press.
- Carruthers, P. (2010) Introspection: Divided and partly eliminated, *Philosophy and Phenomenological Research*, **LXXX**, pp. 76-111. [10.1111/j.1933-1592.2009.00311.x](https://doi.org/10.1111/j.1933-1592.2009.00311.x)
- Carruthers, P. (2011) *The Opacity of Mind: An Integrative Theory of Self-Knowledge*, Oxford: Oxford University Press.
- Cassam, Q. (2014) *Self-Knowledge for Humans*, Oxford: Oxford University Press.
- Chalmers, D. (2003) The content and epistemology of phenomenal belief, in Smith, Q. & Jokic, A. (eds.) *Consciousness: New Philosophical Essays*, Oxford: Oxford University Press.
- Doyle, C. (2019) Review of Alex Byrne's Transparency and Self-Knowledge, *European Journal of Philosophy*, **27**, pp. 515-518. [10.1111/ejop.12459](https://doi.org/10.1111/ejop.12459)
- Dretske, F. (1994) Introspection, *Proceeding of the Aristotelian Society*, **94**, pp. 263-278. [10.1093/aristotelian/94.1.263](https://doi.org/10.1093/aristotelian/94.1.263)
- Dretske, F. (1999) The mind's awareness of itself, *Philosophical Studies*, **95**, pp. 103-124. [10.1023/a:1004515508042](https://doi.org/10.1023/a:1004515508042)

- Dretske, F. (2003) How do you know you are not a zombie?, in Gertler, B. (ed.) *Privileged Access – Philosophical Accounts of Self-Knowledge*, Burlington, VT: Ashgate Publishing Company.
- Dretske, F. (2012) Awareness and authority: Skeptical doubts about self-knowledge, in Smithies, D. & Stoljar, D. (eds.) *Introspection and Consciousness*, Oxford: Oxford University Press.
- Duncan, M. (2021) Acquaintance, *Philosophy Compass*, **16** (3), e12727. [10.1111/phc3.12727](https://doi.org/10.1111/phc3.12727)
- Evans, G. (1982) *The Varieties of Reference*, Oxford: Oxford University Press.
- Evans, J. & Frankish, K. (eds.) (2009) *In Two Minds: Dual Processes and Beyond*, Oxford: Oxford University Press.
- Fernández, J. (2013) *Transparent Minds: A Study of Self-Knowledge*, Oxford: Oxford University Press.
- Gallois, A. (1996) *The World Without, the Mind Within: An Essay on First-Person Authority*, Cambridge: Cambridge University Press.
- Gertler, B. (2001) Introspecting phenomenal states, *Philosophy and Phenomenological Research*, **63**, pp 305-328. [10.1111/j.1933-1592.2001.tb00105.x](https://doi.org/10.1111/j.1933-1592.2001.tb00105.x)
- Gertler, B. (2012) Renewed acquaintance, in Smithies, D. & Stoljar, D. (eds.) *Introspection and Consciousness*, Oxford: Oxford University Press.
- Gertler, B. (2021) Self-knowledge, in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition)  
<<https://plato.stanford.edu/archives/win2021/entries/self-knowledge/>>.
- Goldman, A.I. (2006) *Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading*, Oxford: Oxford University Press.

- Gopnik, A. (1993) How we know our minds: The illusion of first-person knowledge of intentionality, *Behavioral and Brain Sciences*, **16**, pp. 1–14.  
[10.1017/s0140525x00028636](https://doi.org/10.1017/s0140525x00028636)
- Gopnik, A. and H.M. Wellman (1992) Why the child's theory of mind really is a theory, *Mind & Language*, **7**, pp. 145–171. [10.1111/j.1468-0017.1992.tb00202.x](https://doi.org/10.1111/j.1468-0017.1992.tb00202.x)
- Gordon, R. (2007) Ascent routines for propositional attitudes, *Synthese*, **159**, pp. 151-165.  
[10.1007/s11229-007-9202-9](https://doi.org/10.1007/s11229-007-9202-9)
- Hulse, D., Read, C. & Schroeder, T. (2004) The impossibility of conscious desire, *American Philosophical Quarterly*, **41** (1), pp. 73-80.
- Keeling, S. (2020) Review of Alex Byrne's Transparency and Self-Knowledge, *Philosophical Quarterly*, **70**, pp. 280, 639-642. [10.1093/pq/pqz070](https://doi.org/10.1093/pq/pqz070)
- Lawlor, K. (2009) Knowing what one wants, *Philosophy and Phenomenological Research*, **79**, pp. 47-74. [10.1111/j.1933-1592.2009.00266.x](https://doi.org/10.1111/j.1933-1592.2009.00266.x)
- Lycan, W.G. (1996) *Consciousness and Experience*, Cambridge, MA: MIT Press.
- Lycan, W. (2019) Representational theories of consciousness, in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy* (Fall 2019 Edition) <<https://plato.stanford.edu/archives/fall2019/entries/consciousness-representational/>>.
- Moran, R. (2001) *Authority and Estrangement – An Essay on Self-Knowledge*, Princeton, NJ: Princeton University Press.
- Nichols, S. & Stich, S.P. (2003) *Mindreading: An Integrated Account of Pretense, Self-Awareness, and Understanding of Other Minds*, Oxford: Oxford University Press.
- O'Brien, L. (2007) *Self-Knowing Agents*, Oxford: Oxford University Press.
- Oddie, G. (2005) *Value, Reality, and Desire*, New York: Oxford University Press.
- Paul, S.K. (2014) The transparency of mind, *Philosophy Compass*, **9**, pp. 295-303.  
[10.1111/phc3.12126](https://doi.org/10.1111/phc3.12126)

Paul, S.K. (2015) The transparency of intention, *Philosophical Studies*, **172**, pp. 1529–1548.

[10.1007/s11098-014-0363-3](https://doi.org/10.1007/s11098-014-0363-3)

Paul, S.K. (2020) Review of Alex Byrne's Transparency and Self-Knowledge, *Philosophical Review*, **129** (3), pp. 480-484. [10.1215/00318108-8311460](https://doi.org/10.1215/00318108-8311460)

Peacocke, C. (1999) *Being Known*, Oxford: Oxford University Press.

Roche, M. (forthcoming) Review of Alex Byrne's Transparency and Self-Knowledge, *Philosophical Psychology*.

Ryle, G. (1949) *The Concept of Mind*, London: Hutchinson.

Schroeder, T. (2020) Desire, in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy* (Summer 2020)

<<https://plato.stanford.edu/archives/sum2020/entries/desire/>>.

Schwitzgebel, E. (2008) The unreliability of naïve introspection, *Philosophical Review*, **117**, pp. 245–273. [10.1215/00318108-2007-037](https://doi.org/10.1215/00318108-2007-037)

Schwitzgebel, E. (2012) Introspection, what?, in Smithies, D. & Stoljar, D. (eds.) *Introspection and Consciousness*, Oxford: Oxford University Press.

Schwitzgebel, E. (2019) Introspection, in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2019 Edition)

<<https://plato.stanford.edu/archives/win2019/entries/introspection/>>.

Setiya, K. (2012) Transparency and inference, *Proceedings of the Aristotelian Society*, **CXII**, pp. 263-268. [10.1111/j.1467-9264.2012.00334.x](https://doi.org/10.1111/j.1467-9264.2012.00334.x)

Sosa, E. (1999) How to defeat opposition to Moore, *Philosophical Perspectives*, **13**, pp. 141-153. [10.1111/0029-4624.33.s13.7](https://doi.org/10.1111/0029-4624.33.s13.7)

Stampe, D. (1987) The authority of desire, *Philosophical Review*, **96**, pp. 335-381. [10.2307/2185225](https://doi.org/10.2307/2185225)

Stroud, S. & Svirsky, L. (2021) Weakness of will, in Zalta, E. (ed.) *The Stanford Encyclopedia of Philosophy* (Winter 2021 Edition)

<<https://plato.stanford.edu/archives/win2021/entries/weakness-will/>>.

Valaris, M. (2011) Transparency as inference: Reply to Alex Byrne, *Proceedings of the Aristotelian Society*, **CXI**, pp. 319-324. [10.1111/j.1467-9264.2011.00312.x](https://doi.org/10.1111/j.1467-9264.2011.00312.x)

Watson, G. (1975) Free agency, *The Journal of Philosophy*, **72** (8), pp. 205-220.

[10.2307/2024703](https://doi.org/10.2307/2024703)

Williamson, T. (2001) *Knowledge and Its Limits*, Oxford: Oxford University Press.

Zimmerman, A. (2006) Basic self-knowledge: Answering Peacocke's criticisms of constitutivism, *Philosophical Studies*, **128** (2), pp. 337-379. [10.1007/s11098-004-7797-y](https://doi.org/10.1007/s11098-004-7797-y)