Review of Analytic Philosophy
Vol. 1 No. 1 (2021)

Research article

The Transparency of Desire as Motivation

Shuhei Shimamura

Abstract:
Evans claims that we introspect our belief that *p* simply by judging, at the first-order, outward-looking level, that it is the case that *p*. This is sometimes called the transparency account of belief introspection. Recently, several authors such as Moran, Byrne, and Ashwell have attempted to extend similar outward-looking accounts to the realm of attitudes with motivational power, or desires. In this paper, I point out that their attempts fail either in adequately explaining the authority of desire introspection or in being outward-looking. I claim that their failure comes from a common root: their self-ascriptive procedures, unlike Evans's, do not invite us to redeploy the ability to hold the attitude that is being self-ascribed. Instead, I offer a series of alternative transparency accounts of desire introspection satisfying this redeployment condition.

Keywords:
Belief, Desire, Introspection, Self-knowledge, Transparency

1. Introduction

In this paper, I discuss whether and how an account of introspection based on the idea of transparency can be extended to desire. In this section, after briefly looking at why introspection raises a philosophical problem in the first place, I introduce what I take to be a paradigmatic instance of the transparency-based account of introspection, namely, Evans's account of belief introspection. Against this background, I state two main points concerning the transparency of desire that I aim to establish in this paper and outline how I address them in the following sections.

Received September 1, 2020; accepted April 19, 2021; online published May 11, 2021; https://doi.org/10.18494/SAM.RAP.2021.0005

1.1 The problem of attitudinal introspection

It is widely agreed that when we self-ascribe our attitudes, such as belief, intention, and desire, we usually do not have to rely on behavioral evidence. Suppose, for example, that I believe that it is very hot outside. When I self-ascribe this belief by saying, 'I think it is very hot outside', I usually do not have to justify this self-ascription by citing my relevant behavior. Indeed, such a self-ascription often precedes the occurrence of such behavior, such as my carrying iced water with me when I go out.

Even though our attitudinal self-ascriptions seem non-evidential in this way, it is difficult for other people to challenge them. Consider, again, the non-evidentially made self-ascription of my belief that it is very hot outside. Unless there is some special reason to doubt my sincerity, it would be quite difficult, if not impossible, for others to challenge it. In this sense, our non-evidentially made self-ascriptions are authoritative.

But why are our self-ascriptions authoritative in spite of being non-evidential? The key to answering this question seems to lie in the special way we ascribe our own attitudes that is different from the way others ascribe the same attitudes to us.[1] The supposed special method of self-ascription is sometimes called 'introspection'. Then, the question is, what is introspection, and how does it produce non-evidential but authoritative self-ascriptions? Let us call this the problem of attitudinal introspection.[2]

1.2 The transparency account

Different accounts have been proposed to solve the problem of introspection. Traditionally, introspection has been literally understood as the special faculty of an 'inner look'. Contemporary versions of this line of thought have been developed by many authors, such as Armstrong (1993, ch. 15), Lycan (1996, ch. 2), Nichols and Stich (2003, ch. 4), and Goldman

1 However, there are some authors who deny such asymmetry between the first-person and third-person access, such as Ryle (1949, ch. 6), Gopnik (1993), and Carruthers (2011).

2 Following Wright (1998, pp. 13–18), here I introduce the problem of attitudinal introspection by focusing on the distinctive features of linguistic self-ascriptions (what are sometimes called 'avowals'), although the problem also pertains to the corresponding mental self-attributions. Essentially the same problem has been addressed by many other philosophers including Gallois (1996, pp. 15–16), Moran (2001, p. 10), Finkelstein (2003, pp. 100–101), Bar-On (2004, p. 20), Byrne (2005, pp. 80–82), and Fernández (2013, pp. 4–7).

(2006, ch. 9); according to these authors, the faculty of monitoring our inner states is somehow implemented in our brains.[3]

However, closer reflection on how we introspect our attitudes indicates something quite opposite. For example, with respect to the introspection of belief, Gareth Evans writes, 'In making a self-ascription of belief, one's eyes are, so to speak, or occasionally literally, directed outward—upon the world' (Evans 1982, p. 225). Imagine that someone asks us, 'Do you think it is very hot outside?' When we answer this question, our attention seems directed not at our asked belief but rather at the hotness outside. From our introspective perspective, our beliefs seem invisible or transparent, as it were.

Based on this observation of the phenomenological transparency of belief, Evans proposes an interesting account of belief introspection that posits no faculty of an inner glance or monitoring. He writes, 'I get myself in a position to answer the question whether I believe that *p* by putting into operation whatever procedure I have for answering the question whether *p*' (Evans 1982, p. 225).[4] It would be fair to formulate the account suggested here as the following two-step procedure.

(i) Answer the question of 'Is it the case that *p*?'
(ii) If the answer is 'Yes', output, 'I believe that *p*'.

Thus, according to this account, we self-ascribe the belief that it is very hot outside simply by affirmatively answering the question, 'Is it very hot outside?'.

Now, how does Evans's self-ascriptive method let us explain the problem of belief introspection? It is easy to see that his method relies on no behavioral evidence. For, according to it, all we need in order to self-ascribe a belief that *p* is simply answer that *p.* It would be wondered, however, why we are allowed to self-ascribe the belief this way. After all, it is unsound to infer from *p* that someone believes that *p.* Nonetheless, why are we allowed to move from *p* to our belief that *p*? Moreover, why is such a self-ascription even authoritative?

---

3 Note, however, that there are some inner perception theorists who do not undertake this extra commitment of naturalism, such as Gertler (2011) and Paul (2012; 2015).

4 Similar ideas are endorsed by Edgley (1969, p. 90), Gallois (1996, p. 46), Gordon (1996, pp. 15–16), Moran (2001, pp. 60–62), Byrne (2005, p. 95; 2011b, pp. 203–204), Fernández (2013, pp. 40–59), and Valaris (2014, pp. 4–5). Criticisms on such ideas are found in Bar-On (2004, pp. 113–114; pp. 118–122), Goldman (2006, pp. 240–241), Gertler (2010, pp. 191–193; pp. 260–264) (2011), and Carruthers (2011, pp. 81–84). Due to limitations of space, however, I do not get into this debate.

Evans's answer to these questions is not very clear. He claims, 'If a judging subject applies this procedure, then necessarily he will gain knowledge of one of his own mental states: even the most determined sceptic cannot find here a gap in which to insert his knife' (Evans 1982, p. 225); but he does not provide any further support for this claim. Many theorists find his stance unsatisfactory and try to give some more substantial answer (see, e.g., Gallois 1996, p. 47; Moran 2003, p. 405; Setiya 2011, pp. 184–185; Byrne 2005, p. 95; 2011b, pp. 203–204; 2018, pp. 75–77; 99–100; and Valaris 2014).

In my view, one fruitful way to make sense of Evans's claim above is to read him assuming that when we follow his self-ascriptive procedure, we are invited to redeploy our believing ability. It seems that if we sincerely answer that *p*, we usually count as already believing that *p*. In other words, in answering that *p*, we seem to exercise basically the same ability as we exercise to believe that *p*.[5] This justifies the apparently problematic step of Evans's procedure, where we proceed from *our* answer that *p* to the *self*-ascription that *we* believe that *p*. Note that this explanation is valid only in the case of *self*-ascription, where the ascribing subject is identical to the subject of the belief that is being ascribed.

Thus, Evans's account seems to be able to answer the problem of belief introspection. Furthermore, as Alex Byrne (2005, pp. 92; 99) points out, it is an ontologically economical account, since it does not appeal to any extra faculty of introspection. Thus, other things being equal, his account is preferable to those appealing to some special faculty of inner perception.

To sum up, Evans turns the observation of the phenomenological transparency of belief into an economical account of belief introspection. That is, he describes, from the perspective of the introspecting subject, how she self-ascribes a belief not by thinking about the belief itself but by thinking about what the belief is directed at. Let us call such an outward-looking account of introspection the transparency account.

5 Moran seems to suggest a similar point when he writes, 'Any understanding of belief that provides for the minimal idea that believing involves "holding true" will entail that it is at least possible to announce one's belief by reporting on the truth as one sees it.' (Moran 2001, p. 105). For similar points, also see Byrne (2005, pp. 96–98; 2011b, pp. 206–207; 2018, pp. 103–112), Setiya (2011, pp. 186–187), Valaris (2014), and Andreotta (2020). On the other hand, Gertler (2010, pp. 260–264) and Paul (2012, pp. 340–341) caution that even if a subject sincerely answers so and so, she may not count as believing it if she lacks the disposition to act on this belief in relevant subsequent contexts. If so, pace Evans, there may be cases in which Evans's procedure outputs a wrong self-ascription. Note, however, that as long as such mistakes remain exceptional, they are still compatible with the authority of Evans's procedure.

1.3 Two aims

Imagine, for example, that we are asked, 'Do you intend to go the cinema tonight?' Our attention seems directed not at the intention being asked but rather at our going to the cinema. Similarly, if we are asked, 'Do you want to eat cheesecake?', we seem to think rather about our eating cheesecake. Many of our attitudes other than belief also seem phenomenologically transparent. Such an observation is shared by several authors, such as Bar-On (2004, pp. 106–107), Fernández (2007, pp. 524–526; 2013, pp. 87–88), and Byrne (2011a, pp. 174–175; 2018, pp. 158–159). Then, a natural question is, whether and how we can extend the transparency account to those other phenomenologically transparent attitudes.

This paper tackles this question with the focus on desire. I focus on desire because desire, together with belief, seems to be constitutive of rational agency. Extending the transparency account to the introspection of desire would thus be a crucial step toward a transparency-based systematic account of the self-consciousness of a rational agent. Another reason for focusing on desire is that it is closely related to various emotions, if such emotions may not be reducible to desire and belief (see, e.g., Scarantino and de Sousa 2018, pp. 7–8; pp. 29–48; p. 58). Thus, for ambitious transparency theorists who wish to extend their account to a yet wider range of phenomenologically transparent attitudes, desire should be an important touchstone.

Recently, Richard Moran, Alex Byrne, and Lauren Ashwell have suggested their respective transparency accounts of desire introspection (see Moran 2001, pp. 114–116; Byrne 2011a; 2018, pp. 156–166; Ashwell 2013).[6] However, I argue, through §2 to §4, that their accounts fail to maintain the original virtues of Evans's transparency account: they either fail to adequately explain the authority of our desire self-ascriptions or fail to be outward-looking (and thus economical). One of the main aims of this paper is to show that their failure derives from a common root. As I argue in §5, the root is that they miss an essential insight of Evans's transparency account that we self-ascribe a transparent attitude that *p* by redeploying the ability to hold that attitude—namely, by thinking about *p* in the way that amounts to having that attitude.

The other aim of this paper is to offer a schema to construct transparency accounts of desire introspection that respects Evans's insight of redeployment. In §6, I explore a series of procedures for the self-ascriptions of different types of desire that this schema can lead us to. I claim that they have not attracted enough attention in the literature but are promising in economically explaining the non-evidential authority of the introspection of relevant desires.

6 In related contexts, Byrne (2011b) and Setiya (2011) independently propose a similar transparency account of the introspection of an intention, while Barz (2015) offers one for the introspection of a wish.

Before embarking on the discussion, a few points should be clarified. First, I assume that desire is a contentful attitude. Although some desires may appear to be directed at an object rather than a state of affairs (e.g., I want *cheesecake*), I assume that even such desires, if fully specified, turn out to have a content specifiable by a sentence (e.g., I desire *that I eat cheesecake*). Second, it is sometimes debated whether the content of desire is a proposition or something else.[7] In this paper, following orthodoxy, I discuss desire as if it is a propositional attitude. However, nothing crucial hinges on this second assumption.

Finally, in this paper, I use 'desire' as an umbrella term covering the entire range of motivating attitudes, such as an intention, want, hope, and wish. In other words, 'desire' is used here to refer to what Donald Davidson (1963, pp. 3–4) calls pro-attitude. Focusing on desire in this technical sense is a substantial choice and therefore in need of further justification. First, it is a controversial question how our *ordinary* notion of desire should be analyzed (cf. Marks 1986; Schueler 1995, ch. 1; Schroeder 2004, ch. 1). Resolving this controversy is beyond the scope of this paper. Without a clear notion of desire, however, it is difficult to examine discussions concerning the transparency of desire. This trouble can be sidestepped by focusing on the technical notion of pro-attitude, since we have a relatively clear functionalist characterization of pro-attitude (e.g., Smith 1994).

Second, whether or not this technical notion matches our ordinary notion of desire, the former notion has an independent philosophical significance—it plays an essential role in the explanation of our rational agency. As I mentioned above, one of my larger aspirations for extending the transparency account is to obtain a prospect on the systematic account of the self-consciousness of a rational agent. Given this aspiration, it is reasonable to focus on the general category of motivation (i.e., pro-attitude).

My choice to focus on desire as motivation may raise concerns for those who are familiar with the abovementioned work by Moran, Byrne, and Ashwell, since their notions of desire are somewhat different from mine. As it turns out below, however, it can be said that we still share the same subject matter since they all assume that motivation is an essential element of desire. It is just that their target notions of desire are somewhat narrower than mine. In the following, I will be explicit about the differences between our notions when these differences can affect my arguments.

---

7 For example, Castañeda (1974, ch. 2 and ch. 3) classifies the content of intention under the term 'practition', which is distinguished from a truth-evaluable proposition. Barz (2015) invokes a similar idea in his transparency account of the introspection of a wish.

2. Moran and Byrne's desirability view

As we saw above, it seems that our desires are usually phenomenologically transparent. The trouble is, however, that the mere observation of the phenomenological transparency of desire does not determine the exact shape of a transparency account of desire introspection. Certainly, when we self-ascribe the desire to go to the cinema, our attention is directed at its content, our going to the cinema. Yet, exactly what type of question concerning this content do we have to answer to self-ascribe the desire at issue?

Interestingly, Richard Moran and Alex Byrne have reached a similar answer to this question via somewhat different routes. According to them, we self-ascribe a desire that *p* by answering that *p* is desirable.[8] For Moran, this is an application of his rationalist active conception of transparency. His fundamental idea is that a rational agent must have the practical capacity to yield a certain class of attitude simply by virtue of judging that there is a sufficient reason for having that attitude (see, e.g., Moran 2001, p. 131). Given this, a rational agent is in a position to answer whether she has such an attitude simply by answering whether there is a sufficient reason for having it. Moran calls this the 'Transparency Condition' (see, e.g., ibid., p. 84). Applying this general idea to desire, Moran (ibid., p. 115) claims, 'It is the normal expectation of the person, as well as a rational demand made upon him, that the question of what he actually does desire should be dependent in this way on his assessment of the desire and the grounds he has for it.' As to the 'grounds' for desire, Moran says, 'One is an agent with respect to one's attitudes insofar as one … orients oneself toward the question of one's desires by reflecting on *what's worthwhile or diverting or satisfying*.' (ibid., p. 64; emphasis added). Thus, according to Moran, a rational agent self-ascribes her desire that *p* by judging that *p* is desirable.

There are two important points to note about Moran's position. First, Moran limits the scope of his transparency account to what are sometimes called 'motivated' desires (Nagel 1970, pp. 29–30) or 'judgment-sensitive' desires (Scanlon 1998, p. 20). According to Moran, those desires 'depend on certain beliefs about what makes these various things desirable' (2001, p. 115). Moran thus sets aside 'unmotivated' or 'brute' desires that 'simply happen' (ibid., p. 114), such as ones typically associated with hunger or sheer fatigue.[9]

Second, the desirability to which Moran appeals in his account is desirability in the normative sense, not in the explanatory sense. He is explicit about this when he says, 'Rather,

8 Similar ideas are also suggested, albeit in passing, by Boyle (2009, p. 136; n. 25; 2011, p. 237) and Valaris (2014, n. 18).

9 In §2.1, I criticize Moran in this respect.

in conceiving of the questions he's asking as part of the specific activity of deliberating, he has asserted the determination of the answer by *justifying reasons rather than explanatory ones*.' (ibid., p. 132; emphasis added) This choice is important, since otherwise Moran would have to face a difficult task. Taking something desirable in the explanatory sense *is*, by definition, desiring for it. In other words, explanatory desirability is the property that we project onto a thing if and only if we desire it. If Moran appealed to it in his account, he would have to explain how we can judge something has this projected property without recourse to the preceding self-knowledge that we have the desire for it. By appealing to desirability in the normative sense, however, Moran evades this task.

Next, let us turn to Byrne. Unlike Moran, Byrne's starting point is not any general conception of transparency. Rather, he motivates his account of desire introspection by reflecting on what we attend when we self-ascribe particular desires. He invites us to imagine that 'my accommodating companion asks me whether I want to go to the sushi bar across town or the Indian restaurant across the street.' (Byrne 2018, p. 158) Then, he observes, 'I weigh the reasons for the two options—the considerations that count in favor, as Scanlon puts it (1998: 17), of going to either place.' (Byrne 2018, pp. 158–159) After comparing several such cases of desire self-ascriptions, he concludes, 'One's desires tend to line up with one's knowledge of the desirability of the options; that is, known desirable options tend to be desired. (Whether this is contingent, or a constitutive fact about desire or rationality, can, for present purposes, be left unexamined.)' (ibid., p. 161) Consequently, he proposes, 'If φ ing is a desirable option, believe that you want to φ', as his outward-looking procedure of desire self-ascriptions.

Thus, Byrne reaches a transparency account that is similar to Moran's, as it were, from bottom up—from reflections on particular examples of desire self-ascriptions. This is an impressive coincidence. One may wonder, however, to what extent they strictly share the same view. Two points are worth noting. First, as the parenthesized part of the second from the last quotation indicates, Byrne does not commit himself to Moran's rationalist active conception of transparency. Consequently, Byrne does not have to set aside 'brute' desires. Rather, he seems to prefer to talk about desires as we naturally call them (cf., ibid., p. 160). Thus, Byrne's intended explanatory target may be slightly wider than Moran's.

Second, one may wonder if Byrne's key notion of desirability is really the same as that of Moran's, namely desirability in the normative sense. This doubt may seem especially pressing when Byrne writes, 'Going there [i.e., the sushi bar] is … a *desirable* one, in the Oxford English Dictionary sense of having "*the qualities which cause a thing to be desired*: Pleasant, delectable, choice, excellent, goodly."' (ibid., emphasis added) The talk of causing a desire

here may appear to suggest that Byrne's desirability is the explanatory one.[10] However, Byrne explicitly denies this reading in a later section where he addresses what he calls 'the *circularity objection*' (ibid., p. 162), which he describes as the worry that 'in order to find out that something is desirable, one has to have some prior knowledge of one's desires.' (ibid.) There, he contrasts his desirability to desirability in the explanatory sense and characterizes his as 'a reason in the (operative) normative sense, the "consideration in favor" sense of "reason"' (ibid., p. 163). He then argues that desirability in this normative sense does not involve desire in any way that justifies the circularity objection.

At this point, it would be fair to say that Moran and Byrne propose essentially the same transparency account of desire introspection (though they may have somewhat different views about its scope). Let us call this 'the desirability view', which can be formulated as the following procedure.

(Desirability)

(i) Answer the question of 'Is it desirable that *p*?'

(ii) If the answer is 'Yes', output, 'I desire that *p*'.

Recall that what is at issue in (i) is desirability in the normative sense rather than the explanatory sense. Thus, according to this procedure, we self-ascribe the desire to go to the cinema this evening by affirmatively answering, from the evaluative perspective, that it is desirable (worthwhile, excellent, etc.) to go to the cinema this evening.

So far, we have seen how some authors are attracted to what I called the desirability view. Despite its apparent plausibility, there are several serious problems with this view, to which I turn in the following two subsections.

2.1 The problem of limited scope

One immediate problem with the desirability view is its limited scope. We often desire something that we do not consider desirable. For instance, suppose that after eating up a piece of marvelous cheesecake, I crave for another. I envisage, though, that if I have another piece, numerous disagreeable consequences will ensue—I will feel sick, gain weight, blame myself later, and so on. Thus, I consider, from the evaluative viewpoint, that it is *not desirable* to order another piece of cheesecake. Unfortunately, however, my craving for the cheesecake never disappears. When I have such a judgment-insensitive, recalcitrant desire, I can usually self-

10 I thank an anonymous referee for raising this potential worry.

ascribe it immediately. Indeed, such a self-ascription is necessary to put me in a position to self-control that troublesome desire. However, the desirability account cannot cover this important class of desire introspection, where the desired things are not considered utterly desirable.

Moran (2001, pp. 114–115) frankly admits that his account cannot cover all the instances of desire introspection.[11] As mentioned above, Moran sets judgment-insensitive desires outside the target of his rationalist version of the transparency account. To motivate this, Moran points out that judgment-insensitive and judgment-sensitive desires have rather different phenomenologies—he writes, '[Judgment-insensitive desires] may be experienced by the person as feelings that simply come over him' (ibid., p. 114), whereas '[judgment-sensitive desires] may be states of great conceptual complexity, attitudes that we articulate, revise, argue about, and only arrive at after long thought.' (ibid.) According to Moran, transparency (as he understands it) is characteristic of only the latter type of desires.

One problem with Moran's response is, as Sarah K. Paul (2014, p. 302) points out, that it risks a touted advantage of the transparency account—the ontological economy. Given that the transparency account Moran proposes does not apply to judgment-insensitive desires, how does he explain the introspection of those desires? If he appeals to a special faculty, he ends up abandoning the ontological economy. (This is a likely reason for which Byrne, who emphasizes the economical nature of transparency account, is hesitant to commit himself to Moran's rationalist conception of transparency and his consequent bifurcating stance about desires.)

A further problem with Moran's response, I claim, is that it seems ad hoc, since judgment-insensitive desires are often no less phenomenologically transparent than judgment-sensitive desires. When I crave for the cheesecake, for example, my attention seems directed mostly at the cheesecake—its silky surface, smell, taste, etc. I do not deny that the phenomenology of such a judgment-insensitive desire is often different from that of judgment-sensitive ones in other respects as indicated by Moran: Certainly, I experience no deliberating process preceding the craving (rather, the craving 'simply occurs' to me); I also feel some characteristic phenomenal qualities. Yet, such immediacy and qualitativeness are consistent with transparency.

---

11 Another possible reply to this under-generation problem may be as follows: Even if another piece of cheesecake is not desirable *tout court*, it is still desirable in some weaker sense, and it is this latter sense of desirability that is at issue (cf. see Ashwell 2013, n. 10). A challenge with this reply is how to characterize the weaker sense of desirability. One possible way is to focus on *prima facie* desirability, or desirability as *appearing* to us. This line of modification is indeed suggested later by Byrne (2018, p. 166), to which I will return at the end of §2.2. My criticism of it is stated in §3. Another option is to focus on *pro tanto* desirability, or desirability *to some extent*, which is critically examined in §4.

This becomes clearer if we think of another instance of immediate and qualitative attitudes—perceptual beliefs. Suppose that I notice by sight that it is raining. My belief about the rain seems not only immediate (i.e., not based on any prior deliberation) and accompanied by characteristic qualities, but it also seems transparent. Now, it would be strange if a transparency theorist of belief proposed to set aside such a perceptually acquired belief because it is immediate and qualitative. (As far as I know, there is no transparency theorist of belief who proposes this, including Moran himself.) Similarly, transparency theorists of desire should pursue an account that is also applicable to judgment-insensitive desires as long as they are phenomenologically transparent. Moran's stance is ad hoc in that he sets aside such transparent desires without an appropriate reason.[12]

2.2 The problem of limited reliability

Another more serious problem with the desirability view is its limited reliability. For instance, we are sometimes knowledgeably in a state of apathy. Suppose that, after enjoying sea bathing, I am lying exhausted on a couch. I take it as totally desirable to take a shower now—it would be pleasant, relaxing, refreshing, and so on. Nonetheless, I just do not feel like doing so, and I continue to lie idly on the couch. When I am in such an apathetic state, I usually know it, and I do not erroneously self-ascribe the desire to take a shower. Yet, according to the desirability account, I would be allowed to self-ascribe it, since I consider it desirable to take a shower. Here, the desirability account seems to over-generate an erroneous desire self-ascription that we do not tend to make.[13]

---

12 In defense of Moran's bifurcating stance, Boyle (2009, pp. 140–142; pp. 147–156) criticizes the assumption that all introspection should be explained in a uniform manner. I agree with Boyle that uniformity should not be automatically assumed across the board. In fact, in §6 I argue that we should look for different transparency accounts for different types of desires depending on the natures of those desires. It would be worth stressing, however, that my criticism of Moran is not based on the uniformity assumption. Rather, it is based on the observation that judgment-insensitive desires are often phenomenologically transparent. I criticize Moran not for bifurcating desires but for setting aside the introspection of a class of phenomenologically transparent desires without an appropriate reason.

13 One may object that the self-ascriptions based on desirability judgments are never erroneous, since any genuine desirability judgments involve (or are necessarily accompanied by) some (not necessarily dominating) motivations. However, as far as the current case is concerned, such a radically internalist claim seems implausible. If I were motivated to take a shower even a little bit, why would I not take a shower, given that there is nothing preventing me from doing so? (Indeed, lying on the coach with sweaty pants on is rather uncomfortable.) Furthermore, the objection above seems off the point in the first place, since the issue here is not the falsity of the self-ascription but

Two points should be clarified. On one hand, the desirability at issue here is one in the normative sense (taking a shower is considered to be 'pleasant, relaxing, and refreshing'). On the other hand, I nevertheless lack the motivation to take a shower (remember that this is a case of apathy). Since Moran and Byrne both admit that motivation is an essential element of their target types of desire,[14] they have to admit that I lack the relevant desire in such a case. The point is that once we focus on desirability in the normative sense (to avoid the threat of circularity), we can easily imagine a case in which we know that our normative judgment about desirability and our actual desire diverge.

Byrne acknowledges this over-generation problem and attempts to dodge it by modifying his account. According to Byrne (2011a, pp. 177–178; 2018, pp. 164–165), his self-ascriptive procedure—if *p* is desirable, self-ascribe the desire that *p*—is reliable but defeasible, and we know when to defeat it. By scrutinizing the case of apathy, he goes on to specify a complicated defeating condition focusing on a special type of intention characteristic of apathy—an intention to do something (e.g., keep lying on the couch) that seems to be not only incompatible with what is desirable (e.g., to take a shower) but also not desirable in itself. Byrne claims that, when we notice that we have such an apathetic intention, we are in a state of apathy, and therefore, we should defeat the procedure.

One immediate worry with Byrne's reply is whether his modified self-ascriptive procedure is inward looking, since intentions are apparently internal states. Byrne (2011a, p. 183; 2018, p. 167) responds to this concern by referring to his transparency account of intention introspection (see also Byrne 2011b; 2018, pp. 167–172), which is supposed to tell us how to self-ascribe intentions without recourse to a special faculty of introspection. However, Ashwell (2013, p. 253) criticizes this response by pointing out a specific reason to suspect that apathetic intentions are exceptions to his transparency account.

Whether Byrne's inventive treatment of apathetic counterexamples to his procedure works or not, it is not enough to protect his procedure from the general problem of limited reliability. As Ashwell also points out, apathy is not the only case in which we fail to desire what we consider desirable. Suppose, for example, that I am wondering what to do tonight. I consider it desirable to catch up with the current political situations in Japan, but I know I just do not

its presence. The desirability view erroneously predicts that I will self-ascribe a desire that I am hesitant to self-ascribe in reality.

14 Remember that Moran explicitly focuses on *motivated* desires in Nagel's sense. On the other hand, Byrne writes, 'If I concluded that I want to go cycling, I would be wrong. If I really did want to go, why am I still lying on this sofa? It is not that I have a stronger desire to stay put—I couldn't care less, one way or the other' (Byrne 2018, p. 161). In this passage, he clearly assumes that the lack of motivation entails the lack of desires in his more relaxed, everyday sense.

feel like doing so. Instead, I choose to continue to write my ongoing paper, which I also find desirable. This case is not apathetic, as I do what I find desirable, but like the apathetic case above, it is a counterexample to the desirability view. Since I am not apathetic, Byrne's defeating condition does not work here. Thus, his modified procedure still over-generates a wrong self-ascription of desire that I do not make.[15] Of course, if Byrne is inventive enough, he may be able to keep defending his view by coming up with an extra defeating condition whenever a new instance of such known discrepancies between our first-order desirability judgments and self-ascriptions of desires shows up. However, such a response seems rather ad hoc.

Byrne (2018, p. 166) later modified his view by claiming that the intended sense of desirability is the one contrasted to the 'ought-to-be-desired' sense of desirability.[16] An immediate worry is whether this modification is compatible with his previous reply to the circularity concern that the relevant sense of desirability is the normative one. Byrne also hinted that the intended sense is closely related to the appearance of desirability. This further characterization of desirability seems to make his modified view very close to Ashwell's alternative view, discussed in the next section.

3. Ashwell's desirability-appearance view

Scrutinizing the case of apathy, Ashwell concludes that it is difficult to defend the desirability account from the problem of limited reliability. Yet, she suggests that a hope for transparency theorists may still lie in the vicinity of this account. Her key observation is that, in the apathy case, one judges something as desirable without *perceiving* it as desirable; just as one judges that the two lines of the Müller–Lyer illusion are of the same length without *seeing* them as being so. Thus, when I apathetically judge it as desirable to take a shower, it does not appear to me desirable to do so—rather, I somehow reason about the desirability without perceiving it. Based on this observation, Ashwell writes, 'Instead of focusing on *judgments of value*, I suggest transparency theorists ought to be looking at something closer to *appearances of value*' (Ashwell 2013, p. 254; emphasis in original).

Since Ashwell offers no explicit procedure of desire self-ascriptions, it is not entirely clear how the proponents of the desirability view ought to modify their procedure according to her suggestion. The closest she offers to such a procedure is as follows: 'Wanting things makes

15 See Ashwell (2013, p. 253) for other non-apathetic counterexamples.

16 This modification was not made in Byrne (2011a), where he originally presented his desirability view. He added it after he had received Ashwell's criticisms mentioned above.

you see them in a certain light, and this is how you introspectively know what you want' (Ashwell 2013, p. 255). A modified procedure naturally suggested by this passage would be as follows.

(Desirability Appearance)
(i) Answer the question of 'Does it appear desirable to me that *p*?';
(ii) If the answer is 'Yes', output 'I desire that *p*'.

Let us call this the desirability-appearance view. Assuming the alleged intimate connection between appearances of desirability and desires, this account seems immune from the problem of limited reliability.

3.1 The problems of limited scope and inward-lookingness

Before critically examining the desirability-appearance view, to be fair, let me mention that Ashwell herself addresses a possible concern of it. She admits that her position may compromise the touted ontological economy of the transparency account by positing an extra faculty for value-perception (i.e., a faculty of seeing good) as distinguished from the one for desire (Ashwell 2013, p. 255). She is ambiguous about to what extent transparency theorists should be satisfied with her view.

However, there are at least two more problems with the desirability-appearance view. One problem is again concerned with its scope. There is a reason to doubt that the scope of this alternative view is wide enough to cover the introspection of transparent desires across the board. As Stocker (1979, pp. 746–749) and Velleman (1992, p. 17) argue, in some occasional moods, we may desire something exactly because it appears *simply bad and of no value at all*. Stocker describes one such mood as follows: 'The whole day has gone so badly, I might as well complete it by ruining the little I did accomplish' (p. 748). Unlike pathological desires, such a perversive desire normally seems transparent and perfectly introspectable. However, it cannot be self-ascribed via the desirability-appearance procedure, since its content, ex hypothesi, does *not* appear desirable to me.[17]

[17] Opposing this last point, one may argue that since I am being attracted to ruining the little I have accomplished, doing so must appear attractive to me. (I thank an anonymous referee for raising this possible objection.) This objection relies on the general assumption that if we desire that *p*, *p* must appear desirable to us. However, it is this assumption—which Velleman calls the 'guise of good' view—that I am challenging here, following Velleman, by referring to Stocker's case. If Stocker is right, it is possible—and indeed is sometimes the case—for us to come to desire that *p* just because

Another more serious problem is that the procedure at issue is apparently inward looking. After all, *the appearance* of desirability, as contrasted with the property of desirability itself, is a mental property—something that occurs to our mind. Thus, the desirability-appearance view just replaces the original self-ascriptive question (i.e., 'Do I *desire* that *p*?') with another self-ascriptive question (i.e., 'Does it *appear* desirable *to me* that *p*?').[18] Of course, an extra story might be told about how the latter question is answered via some outward-looking considerations. Unfortunately, however, Ashwell offers no such story. Until such a story is provided, it is not clear whether the desirability-appearance view sheds any substantive light on our original question of how we introspect our desires.

4. The pro-tanto-desirability view

To finish our critical examination of desirability-focusing approaches to the transparency of desire, let us turn to yet another possible variant of the desirability view that has not been carefully examined in the existing literature. Given the above criticisms of the desirability and desirability-appearance views, this may be a view that people who are still sympathetic to the desirability-focusing approach want to pursue. Below, first I explain why this view may appear attractive at first glance. Then, in the next subsection, I point out a serious problem with this view.

Both of the views examined so far implicitly assume that in introspecting desires, we always attend to a unique value-oriented question. However, this assumption may be challenged based on the following observation. When we ask, 'Do I desire such and such?', our attention seems to wander among different evaluative aspects of the intentional object in question—for instance, 'Is this hamburger tasty?', 'Is it large enough to satisfy my hunger?', 'Is it cheap?', and so on. This further observation may lead us to give up the above assumption that there is a *unique* evaluative question that we answer *whenever* we self-ascribe a desire. Instead, we may want to retreat to a logically weaker claim: *Whenever* we self-ascribe a desire, there is

p appears harmful and of no value at all. I do not see why, in this perversive case, the mere fact of *p* being desired by us would have to add any appearance of desirability to *p*. Rather, *p* must continue to appear valueless, since that is why we came to desire it and continue to do so. Otherwise, we would stop desiring it. Since such a perversive case seems coherently imaginable, I claim that the onus is on those who deny this intuition.

18 This concern should not be confused with Ashwell's concern mentioned above. My concern is rather that, to initiate the kind of self-ascriptive procedure that Ashwell seems to suggest, it is not sufficient that we simply see value; it is also required that we are aware of ourselves seeing value. (Otherwise, we cannot answer 'Yes' to 'Does it appear desirable to me that *p*?'). The question is how such introspective awareness of value-perceptual states is obtained in the first place.

some evaluative question that we answer, where the question may differ depending on who self-ascribes the desire, when she does so, what content the desire has, and so on. Indeed, this can be seen as another way to understand Byrne's modified weaker sense of desirability mentioned at the end of §2.2.

Following this line of thought, we are led to what I call the pro-tanto-desirability view. According to this view, we self-ascribe a desire by favorably evaluating the intentional object of the desire in at least *some respect*.

(Pro Tanto Desirability)
(i) For some x, answer the question of 'Is it desirable in x that *p*?';
(ii) If the answer is 'Yes', output 'I desire that *p*.'

By this procedure, we self-ascribe our desires by evaluating various aspects of the desired things: for example, 'The cinema is fun (i.e., desirable in the recreational viewpoint). Thus, I want to go to the cinema'; 'The cheesecake is delicious (i.e., desirable in taste). Thus, I want to have it'; 'The hamburger is very big (i.e., desirable in size). Thus, I want to eat it', and so on.

This view appears to have two merits over the former two views. First, since it is based on a logically much weaker claim than the desirability view, it is difficult to refute it by raising a counterexample. To do so, we must offer a case in which the subject introspectively self-ascribes a desire while attributing *no* pro tanto value to what is desired. Second, in contrast to the value-appearance question, answering a pro-tanto-value question (e.g., 'Is the hamburger big enough?') does not seem to be a matter of introspection. Therefore, there is no apparent concern about violating the outward-looking condition.

4.1 The problem of unreliability

So far, so good. The pro-tanto-desirability view circumvents the problems that are stumbling blocks for the other two. Still, it has its own serious drawback, since it leaves an essential issue utterly unexplained—that is, why the self-ascriptive method it describes is reliable. According to the view, we self-ascribe our desires by answering different pro-tanto-value questions under different circumstances. Yet, how do we determine which question to answer in order to self-ascribe a given desire under a given situation?

If we take the pro-tanto-desirability view seriously, this issue of value choice is essential for understanding our authority in introspecting desires. For if we are careful and fair-minded,

we can find some pro tanto values in almost all things, including those for which we have no desire at all. Consider, for example, a Beckerian economist trying to understand why kidnapping is rampant in certain countries. Analyzing relevant data, she may conclude that in those countries the expected profit of kidnapping massively outweighs its expected cost—thus, kidnapping is desirable there from a purely economic viewpoint. The point is that she can draw this conclusion even if she has no desire to kidnap. Or if you prefer a mundane example, choose whatever food you hate, and try to find some valuable aspects of it. I do not even want to see cherry tomatoes, but that does not prevent me from admitting that they are healthy, helpful for burning unnecessary fat, and so on. We seem to have the capacity of rational freedom to form some pro-tanto-desirability judgments while bracketing our own desires.[19] Thus, if we self-ascribed desires simply by making *any* pro-tanto-value judgments, we would make many more mistakes than we actually do.

Then, how do we choose, from a bunch of irrelevant ones, such pro-tanto-value questions that answering them allows us to authoritatively self-ascribe a desire in question? In other words, how do we distinguish those pro tanto values that our desires are currently responsive to from those that they are not? The view under consideration offers no account of this issue. Furthermore, it is not clear how the view could offer such an account without smuggling in the introspection of desire at some stage.

5. Where do Moran, Byrne, and Ashwell go wrong?

In the previous sections, we have critically examined three attempts to extend the transparency account to desire introspection. Although the problems with these views may appear diverse and mutually independent, I claim that they come from a common root. In this section, I attempt to pin down the root. Based on the result, in the next section, I suggest a direction that transparency theorists of desire introspection should explore.

---

19 One may still suspect that whenever we find something desirable in some respect, we must have a desire for it; but it is just that the strength of that desire may be very weak (I thank an anonymous referee for raising this point). On one hand, I agree with this suspicion if pro tanto desirability is understood in the explanatory sense—after all, finding something desirable in some respect in the explanatory sense *is* desiring it to some extent. However, as I pointed out in §2, if one appeals to desirability in this projected explanatory sense, one needs to explain how we come to know that a thing has that projected property without recourse to prior self-knowledge of our relevant desire. Otherwise, the account would be circular. On the other hand, I think the above suspicion is implausible if pro tanto desirability is understood in the normative sense. The cases offered above are apparent counterexamples. Furthermore, the case of apathy from §2.2 also counts as a counterexample (given that desirability simpliciter is logically stronger than pro tanto desirability).

In order to specify at which point the views above go wrong, let us briefly review how we have been led to them. The common starting point of transparency accounts is the observation of phenomenological transparency. When we are introspectively aware of a desire, our attention rather seems to be directed at the thing desired. According to transparency theorists, this observation sheds light on how we introspect our desire. That is, we introspect a desire by thinking something outward-lookingly about the thing desired. The question is then exactly what outward-looking thought this is.

Moran and Byrne independently observed that when we wonder whether we desire that *p*, we often wonder whether *p* is desirable. Thus, they proceeded to claim that it is by thinking this thought about desirability that we come to self-ascribe our relevant desire. Natural as this transition appears to be, I claim that this is the step at which they go wrong. Recall how the two problems with Moran and Byrne's desirability view arise. The problem of scope arises because we sometimes desire things without judging them to be desirable. The problem of reliability arises because we sometimes judge things to be desirable without desiring them. Although desires and desirability beliefs often go together (as they should), they are distinct mental states. This is the root of the troubles with the desirability view.

According to Ashwell's desirability-appearance view, we self-ascribe a desire that *p* rather by considering that *p* appears desirable. However, since *p*'s appearing desirable is itself our state of mind, thinking this seems to require the use of introspection. Thus, the problem of inward-lookingness arises. Notice that here again the problem arises from the fact that the focal thought at issue—this time, the second-order thought about desirability appearance—is not equitable with the target attitude of desire.

Finally, the problem with the pro-tanto-desirability view can also be traced back to the same root. The gist of the problem was that since we often admit that *p* is desirable in some particular respect without desiring that *p*, merely thinking the former thought does not guarantee the existence of the latter desire.

To sum up, all three views above focus on, from the perspective of the introspecting subject, what she thinks about the thing desired when she introspects her desire. I believe that this is perfectly in line with the spirit of the transparency account. Why do they nonetheless fail to explain the non-evidential authority of desire introspection? As we have seen above, the root of all the indicated problems can be traced back to the fact that none of the thoughts focused on by these views is equitable with desire. Therefore, unlike Evans's procedure, none of the

procedures suggested by these views is one in which we are invited to redeploy the ability to hold the relevant desire. In my diagnosis, this is where Moran, Byrne, and Ashwell go wrong.[20]

6. The transparency of desire as motivation

The above diagnosis of where the existing views in the literature fail gives us a general condition for a transparency account to succeed. A successful transparency account of the introspection of attitude ψ must be one in which we are invited to exercise the ability to ψ. To specify such a procedure, we must focus on the thought in which ψ consists. It is this thought that must be described from the outward-looking perspective of the introspecting subject. In other words, we must describe, from our outward-looking perspective, what we think about *p* when we ψ that *p*.

As an illustration of such a transparency account, it is helpful to review Evans's account of belief introspection. Recall that according to Evans, we self-ascribe a belief that *p* by thinking that *p* is the case. That *p* is the case is how we describe, from our perspective, what we think about *p* when we believe that *p*. Thus, Evans's account is a typical instance of the successful schema of the transparency account outlined above.

---

20 One may suspect that this criticism does not apply to Moran for the following reason. As we saw in §2, Moran argues that judging *p* to be desirable is all that is needed for a rational agent to form the desire that *p*. If so, it may seem that the ability to form desirability judgments is identical with the ability to form relevant desires. Then, Moran's desirability-based self-ascriptive procedure may appear to satisfy the redeployment condition (cf., Boyle 2011, §II). However, this appearance is misguided, since as the examples from §2.1 and §2.2 indicate, it is possible for us to exercise one ability without exercising the other. They are thus distinct abilities, and they coincide only when we are ideally rational. As an ironical consequence, Moran's procedure can only explain the self-knowledge of those desires that are formed in the rational, judgment-sensitive manner—desires that are in no need of special attention (see §2.1).

To be fair, however, let me mention that Moran is willing to accept this consequence. He highlights the above type of self-knowledge as being constitutive of rational agency in a broadly Kantian sense, while refusing to assimilate it to Cartesian self-knowledge whose supposed function is 'supervision'. According to Moran, transparency is the distinctive feature of such Kantian self-knowledge. (Moran 2001, pp. 107–120; p. 127; pp. 138–151; also see Boyle 2009 for further elaboration of this point). Now, I do not deny that it may be of some deep philosophical importance to distinguish Kantian and Cartesian self-knowledge. I only disagree with Moran about the very last point—about whether transparency is characteristic of Kantian self-knowledge alone. After all, as I pointed out in §2.1, not only judgment-sensitive desires but also judgment-insensitive, recalcitrant desires are often phenomenologically transparent. This gives us a prima facie reason to pursue a transparency account that can cover the self-knowledge of such not ideally rational desires. This is what I attempt to do in the next section.

In this last section, I explore possible transparency accounts of desire introspection within this successful schema. Our main task is to describe, from our perspective, what we think about *p* when we desire that *p*.[21] To address this task, first we need to clarify the nature of desire at issue here.[22] Remember that my notion of desire is pro-attitude. An important advantage of focusing on pro-attitude is, as mentioned in §1.3, that it is a relatively clear functional notion. Pro-attitude is essentially motivation. When we have a pro-attitude for *p*, we are motivated to bring about *p*. In the most primitive case, such a motivation simply moves us to bring about *p*. However, the functional connection between a motivation and an action is not always that straightforward. One possible factor that deters us from bringing about *p* is other competing motivations. Another factor is the belief that there is no available means to bring about *p*. Motivation is thus characterized as a functional state that disposes us to bring about *p* in the absence of these deterring factors. That is,

> (Motivations)
>
> We are motivated for *p* if and only if (we are disposed to make it the case that *p*, unless either (i) we are more strongly motivated for some *q* that we believe is incompatible with our making it the case that *p*, or (ii) we believe that *p* is infeasible).[23]

Given this functional characterization, we can classify motivations according to how closely they are connected to relevant actions. The closest ones, which are described above as the most primitive case, roughly correspond to what we call intention. At the opposite end of this spectrum are what we call a hope or wish. Many of what we usually call desires are located in the middle. To specify transparency accounts of these three classes of motivation, below I attempt to describe, from our outward-looking viewpoint, how we think about *p* when we are motivated for *p* in the different manners described just above. In §6.1, I begin by focusing on the first class of motivation, which roughly corresponds to intention. Here, *p* is roughly described, from our outward-looking viewpoint, as something that we will bring about. Since

21 For those who are sympathetic to desirability-focusing approaches to the transparency of desire, this task can be paraphrased as follows: the task is to describe, from our perspective, what we think about *p* when *p* is *desirable for us in the projected explanatory sense*. From this angle, my approach may be seen as yet another variant of desirability-focusing approaches.

22 Andreotta (2020, §4.2; §6) stresses a similar point, although he proposes transparency accounts of a desire and wish that are different from the ones proposed below. The differences come from the different views we have on the nature of our target attitudes.

23 Note that this is not a reductive, behaviorist analysis of a desire as a motivation in terms of dispositions to act, since the right-hand side of (Motivations) refers to other motivations and beliefs. Rather, it is a functional characterization.

pro-attitude is a functional state that disposes us to bring about something, I take this outward-looking description to give us the natural base for exploring transparency accounts of the other types of pro-attitude as well. However, as we may sometimes be deterred from pursuing what is motivated for, we need to appropriately weaken the above description to obtain transparency accounts of such deterred motivations. In §6.2, I weaken the description to cover motivations defeated by incompatible motivations. In §6.3, I further weaken it to cover motivations that are taken to be infeasible.

Although I suggested above that my three-layer classifications of motivation according to the closeness to action roughly correspond to an intention, desire, and hope or wish, I do not intend to claim that I have offered an analysis of any of our ordinary notions of these conative attitudes. As I mentioned in §1.3, giving such an analysis is a controversial task. In the following subsections, I shall be explicit at places where possible mismatches between my functional classifications and ordinary notions can raise an issue, and I will suggest how such issues can be addressed.

6.1 Predominant motivations

Imagine, again, that I want to go to the cinema this evening. As I mentioned above, the motivation at issue here can be distinguished according to how closely it is connected to the relevant action. Let us begin with the closest case, where I am not deterred from pursuing my motivation. Now, the crucial task is to describe, from the outward-looking perspective of me as the motivated subject, how I think about going to the cinema. According to the successful schema of the transparency account, such a description reveals the way I outward-lookingly self-ascribe the motivation at issue.

At first glance, an obvious answer to the crucial task above is as follows. When I am not deterred from going to the cinema, I simply think that I will go to the cinema. This observation may seem to suggest the following outward-looking procedure to self-ascribe the motivation at issue: I self-ascribe the motivation to go to the cinema simply by affirmatively answering the question of whether I will go to the cinema.

However, this conclusion is too quick. For we sometimes just anticipate that we will do something without self-ascribing the motivation to do it.[24] For example, suppose that getting up at 7 a.m. is my long-term habit. I am so accustomed to this habit that I automatically get up

24 A similar point was originally made by Jonathan Way (2007, p. 225) in his discussion about the transparency of intention. Byrne (2011, p. 217) rediscovered it as a possible problem for his transparency account of intention. Also see Andreotta (2020).

at 7 a.m. every morning. Then, it is quite reasonable for me to anticipate that I will get up at 7 a.m. tomorrow again, even if I do not take myself to particularly want this to be the case. In both the cinema and awakening cases, I think I will do a certain thing, but in the former case, I take the relevant thing as what I am motivated to do, while in the latter I do not. To understand how we distinguish our motivation from mere anticipation, we need to describe, from our outward-looking perspective, the thought involved in the motivation at issue more specifically.

We can get a prospect of fulfilling this task of further specification by turning to the observation independently offered by Byrne (2011b, p. 220; 2018, pp. 171–172), Setiya (2011, p. 193), and Shimamura (2011, p. 149). Inspired by Anscombe's (1963, pp. 13–15) famous remark on the non-observational self-knowledge of intentional action, they observe that when we intend to φ, we think that we will φ, but we also think that this thought does not depend on relevant evidence. Although they focus on intentions instead of predominant feasible motivations, the same insight can clearly be adopted for our current purpose. That is, our target thought—the thought that amounts to a predominant motivation instead of a mere anticipation—can be described, from our outward-looking perspective, as non-evidential in this Anscombean sense: 'I will φ, *whether or not there is evidence that I will*'. The added non-evidential clause is supposed to rule out mere anticipation, since anticipation usually relies on evidence, such as my repeated experience of getting up at 7 a.m.

Thus, we are led to the following procedure for the self-ascription of predominant feasible motivations.

(Predominant Feasible Motivations)

(i) Answer the question of 'Will I make it the case that *p*, whether or not there is evidence that I will?'

(ii) If the answer is 'Yes', then output 'I desire that *p*'.

According to this procedure, I self-ascribe my motivation to go to the cinema by affirmatively answering that I will go to the cinema, even though there is not yet any available evidence that I will do so.

Before addressing possible criticisms of this procedure, let me mention one side point. The above procedure is similar to the ones proposed by Setiya (2011, pp. 192–197) and Byrne (2011b, pp. 217–220; 2018, pp. 168–172) as transparency accounts of the introspection of

intention.[25,26] In my view, however, whether the above procedure counts as an account of intention introspection depends on the controversial topic of how we should understand the nature of intention (cf. Paul 2014, pp. 328–329). If one simply identifies intention with the predominant feasible motivation (cf. Davidson 1963; Sinhababu 2013), then one should agree with Byrne and Setiya. However, if one believes that intention is more than such motivation (cf., Bratman 1987), then one should not be satisfied with Setiya and Byrne's account. For those who endorse the latter position, an interesting question is whether and how the above procedure can be elaborated into the right transparency account of intention introspection. The answer depends on the details of the functional analysis of intention, which is beyond the scope of this paper.

Now, there are two possible criticisms of the above procedure.[27] First, Byrne might criticize that the procedure is defective as an account of the self-ascription of *desire*, since there is a counterexample to it. Suppose, for example, that at midnight I feel too tired to study, but if I do not study, I am sure that I will fail an exam tomorrow. So I make up my mind to study. If asked at this point, a natural thing for me to say is, 'I will study, though I do not want to.' Byrne might thus claim that my procedure over-generates a false self-ascription of desire that we do not actually tend to make.

This is a place at which it matters which particular sense of 'desire' is under discussion. I claim that the above criticism is misguided, since it appeals to a different sense of 'desire' from the one I use. Remember that in this paper I use 'desire' in the technical sense of pro-attitude,

---

25 One subtle difference between Byrne's view and mine is, however, that Byrne (2011, pp. 218–219) rather regards the Anscombean observation as letting him specify a defeating condition of his transparency procedure of intention self-ascriptions, as follows. If you answer 'I will φ', output 'I intend to φ'; *however, if your answer rests on evidence, defeat the output*. Byrne admits that since having evidence or not is a matter of belief, the outward-lookingness of this defeating condition depends on the transparency of beliefs. However, Baker (2015, pp. 3046–3048) argues that even if beliefs are transparent, Byrne's defeating condition turns out to be inward-looking. Baker points out that his defeating condition concerns not just one's belief about evidence but also the supporting relation between the evidential belief and one's answer to the focal question above. According to Baker, there is no way to explain the introspection of such a relation in terms of transparency, since the relation involves causal (or some other counterfactual) dependence between these two attitudes. Now, I am not totally convinced by this last claim of Baker. In any case, however, one advantage of my procedure over Byrne's is that mine is immune to Baker's criticism, since it, unlike Byrne's, does not refer to any supporting (or resting) relation between attitudes.

26 There are some authors who claim that such an Anscombean condition is not enough to rule out mere anticipations, such as Boyle (2011, p. 234) and Paul (2014, pp. 301–302; 2015, pp. 1534–1535). Although I think that this is a serious challenge, I do not get into it here due to limitations of space.

27 I thank the anonymous referees for pointing out these potential challenges below.

or motivation in general. The above example is not a counterexample to my self-ascriptive procedure of predominant motivation, since there I *am* predominantly motivated to study, and, if asked, I will admit that I am so motivated—I will *not* say, 'I sit in front of my desk instead of going to the bedroom. But I don't know why.'

Having said this, I agree with Byrne that there is a perfectly natural sense of 'want' and its synonyms that is narrower than mine, as illustrated in the example above. My current procedure is not fine-tuned to distinguish desires in this narrower sense from motivations in general. Whether and how such fine tuning can be done depend on one's view on the nature of desire in the sense at issue. As mentioned in §1.3, however, this is a controversial topic, and Byrne does not explicitly state his own view on it.[28] Until it becomes clear exactly what Byrne means by 'desire' in the ordinary sense, it is difficult to further develop an answer to his possible challenge of fine tuning.[29]

Another possible concern about the account above is whether its focal question, 'Will I make it the case that *p*, whether or not there is evidence that I will?', is genuinely outward-looking. One may criticize that it is inward-looking since answering it requires us to rely on the introspection of our predominant motivation to bring about *p*—that is, since we think, 'I am predominantly motivated to bring about *p*. *Therefore*, I will make it the case that *p*.'[30] In my view, however, this criticism stands upside down. Remember how we have been led to the focal question above. We have attempted to describe, from the motivated subject's perspective, how she thinks about *p* when she is predominantly motivated to bring about *p*. If we succeed in this attempt, the thought we have described is the motivation itself—more precisely, the

---

28 Byrne (2018, p. 160; n 10) suggests that 'desire' in the ordinary sense refers to something more than mere motivation, but he is not very clear about what that something extra is.

29 What if, one may still wonder, desires in the narrow sense (call them 'desires proper') may not be fully characterizable solely in terms of the way their intentional objects are thought about? For instance, it may be essential for my desire to sleep to count as a desire proper that it tends to yield pleasure when satisfied. In addition, the introspection of a hedonistic tone may be beyond the reach of any transparency account, since such a tone is something that we can immediately feel—something that is accessible without a detour to any outward-looking thinking. My response to this potential challenge is that if desires proper essentially involve non-transparent hedonic tones, there is no reason to pursue a pure transparency account for the introspection of those desires. Note, however, that transparency theorists are allowed to supplement their account with whatever extra account that is apt for explaining the introspection of the non-transparent phenomenological feature(s) of desire proper, unless those theorists commit themselves to the uniformity assumption that the introspection of any desires whatsoever should be explained in a unique outward-looking manner.

30 A similar view is suggested by Grice (1971, pp. 15–19), although he uses his regimented notion of willing instead of motivation.

motivation described from the perspective of the motivated subject. Thus, my non-evidentially maintained thought that I will bring about *p* is what my motivation to bring about *p* consists of, not a consequence from it.

However, one may still wonder how we can know that we will bring about *p* in the absence of relevant evidence. In response to this remaining concern, two points should be clarified. In the first place, it is controversial whether and in what sense we *know* that we will bring about *p* when we say so in the Anscombean non-evidential manner. When I answer, 'I will go to the cinema', it seems weird to challenge me by asking, 'How do you know that? Show me evidence.' It seems to follow that if I ever know that I will go to the cinema, it is not the ordinary type of knowledge, for which it is legitimate to ask for evidence. Rather, it seems to be what Anscombe notoriously calls 'practical knowledge'.[31]

Second, although it is controversial whether practical knowledge counts as genuine knowledge, the debate does not affect my purpose here. My claim is simply that we self-ascribe our predominant motivation to bring about *p* by non-evidentially answering, 'I will bring about *p*'. Even if this answer fails to express knowledge, that does not threaten the reliability of my procedure. Consider the parallel situation in the case of belief self-ascription, where we answer, 'It is the case that *p*', without actually knowing that *p* (either because *p* is false, or because we lack a justification for *p*). Even in such a situation, as Byrne (2018, p. 107) correctly observes, we can reliably self-ascribe our belief that *p* via answering that *p*, since if we answer so, we already believe so. Analogously, if we non-evidentially answer, 'I will bring about *p*', we are already motivated to bring about *p*. Thus, my procedure is reliable, whether or not the non-evidential answer counts as knowledge.

6.2 Defeasible motivations

So far, I have sketched how the schema of a successful transparency account can be applied to the introspection of desire in the primitive case where we are predominantly motivated to bring about things that we think are feasible. However, this cannot be the end of the story. Unlike primitive animals, we do not always try to bring about things we are motivated for even when we believe they are feasible. Rather, we often refrain from doing so in favor of something else that we believe is incompatible with it. For example, even though I want to go to the cinema tonight, I will not try to do so if I have work to be done by today. In such cases, however, our motivations are usually transparent and introspectable. Therefore, a criticism of my procedure above is that its scope is limited.

31 For a criticism of this view, see Grice (1971, pp. 14–15).

The next challenge is how can we extend the scope of the above procedure to the cases of introspectable motivations that are currently defeated? According to our general schema, this question boils down to the question of how we think about bringing about *p* when we are motivated to bring about *p*, but we are not trying to pursue it in favor of something else. For example, when I am motivated to go to the cinema but not trying do so due to work to be done, how do I think about going to the cinema?

A natural thing for me to think would be something like this: 'I would go to the cinema tonight if it were not for work to be done.' Or if we use the ceteris paribus clause to set aside conflicting considerations,[32] the thought can also be described as follows: 'Other things being equal, I would go to the cinema tonight.' This observation, together with the above observation of non-evidentiality, leads us to the following self-ascriptive procedure.

(Defeasible Motivations)

(1) Answer the question of 'Other things being equal, would I make it the case that *p*, whether or not there is evidence that I would?'; and

(2) If the answer is positive, output 'I desire that *p*.'

In §6.1, I pointed out that when we are motivated to bring about *p*, *p* is described, from the perspective of the motivated subject, as something to be brought about by her. In the current case, however, since *p* is not the foremost thing for her to bring about, *p* is, from her perspective, not simply something that she will bring about. To become aware that she is motivated for *p*—i.e., that *p* is what she *would* bring about—, she needs to subjunctively set aside things that may conflict with her pursuing *p*. This is what the ceteris paribus clause in the focal question above does. By thus weakening the previous focal question with the ceteris paribus clause, the current procedure extends the scope of the previous one to defeated desires. A positive answer to the weakened focal question may vary in its degree of hesitation. If the answer is 'Absolutely', we self-ascribe the corresponding desire while granting it the top priority. However, even if the answer is a more hesitant one, such as, 'Yes, other things being equal, I would', we still self-ascribe the corresponding desire. We may occasionally be more specific about what the ceteris paribus clause is ruling out, such as 'If it were not for work to be done, I would'. In such a case, we are also aware that the self-ascribed desire is outweighed by certain conflicting consideration(s).

32 For a similar use of the ceteris paribus clause in the analysis of desire, see, e.g., Hampshire (1975, p. 36).

6.3 Infeasible motivations

It may still be argued that the scope of our second procedure is still limited because the procedure lets us self-ascribe only those motivations that we regard as feasible. However, we sometimes introspect that we are motivated for things that seem to us very difficult or even impossible to bring about. Such motivations seem to roughly correspond to what we sometimes call hopes or wishes. For example, imagine that a hurricane hits a town in which a friend of mine lives. Upon hearing this news, I murmur to myself, 'I hope he is safe'. Yet, it is obvious to me that I am not in a position to make this happen. Thus, I cannot self-ascribe this hope even via my extended procedure. A similar problem can also be raised with respect to the introspection of a wish by slightly changing the situation (e.g., I already know that my friend was injured by the hurricane).[33]

Again, the key to answering this challenge lies in the reflection on how we think about what we hope (or wish) for, when we hope (or wish) for it. Remember that what I call a hope (or wish) here is motivation for something that we believe is unlikely (or impossible) to be brought about. The functional connection of a hope (or wish) to action is very weak. To excavate such a weak type of motivation from within the perspective of the motivated subject, the subject needs to rely heavily on the ability to subjunctively set aside deterring factors—this time, she needs to subjunctively set aside relevant aspects of reality that seem to prevent her from bringing about *p*. In the above case, I need to ask, 'If I were in a position to keep my friend safe, would I choose to do so?' A positive answer to this question seems to imply that I hope for my friend's safety.[34] This observation, together with the observations from §6.1 and §6.2, jointly suggests the following self-ascriptive procedure for hope.

(Infeasible Motivations)

(i) Answer the question of 'If I were in a position to choose whether *p*, other things being equal, would I choose that *p*, whether or not there is evidence that I would?'

(ii) If the answer is positive, output 'I hope that *p*'.

If *p* is obviously incompatible with what we believe can be the case, this procedure may rather be used for self-ascribing a wish.

---

33 For similar points, see, e.g., Schroeder (2004, pp. 16–18), Paul (2014, p. 302; 2015, pp. 1534–1535), and Samoilova (2016, pp. 3374–3375).

34 A similar observation is made by Shoemaker (1996, p. 237), though not in the context of the transparency account.

Two points should be noted with respect to this heavily subjunctive procedure for self-ascribing a hope or wish. First, I recognize that due to its subjunctive character, the focal question above is sometimes rather difficult to answer. I claim, however, that this is a feature rather than a bug. In my view, we do sometimes answer such a difficult question and self-ascribe the corresponding hope or wish that is buried deeply in our mind. Consider, for example, the question of 'If you were able to go back to the past, what would you do?' It will take a fair bit of imagination to answer such a question, but if I come to answer, for example, 'Well, I would tell my late father that he should immediately go to a large hospital before his disease gets too serious', it seems natural for me to self-ascribe the wish to have done so. In my view, excavating our own hope or wish by asking a subjunctive question is rather a prevalent practice. (Consider, e.g., 'If you won a million dollars, what would you buy?', 'If you were in a position to date a celebrity, who would you choose?' etc.)

Second, I also admit that we sometimes make a mistake in answering a relevant subjunctive question. Suppose, for example, that I am talking on the phone with Ken's wife about his non-life-threatening but agonizing disease only curable by a special operation that costs a million dollars. Sympathizing with his agony, I sincerely say to her, 'If there was anything I could do for him, I would definitely do it.' However, such a statement, even if sincere, may turn out to be false. Just after I hang up the phone, I happen to notice on TV that I have won a million dollars in the lottery. After recovering from the first shock of this surprising news, I start to wonder what I should do with that money. At this point, I may realize that despite what I have just said to Ken's wife, I am not willing to use it for him (I murmur to myself, 'After all, he is not dying. I have more important things to do with the money.') It would be natural to conclude from this that I did not predominantly wish to save Ken, even though I sincerely said that I would help him if I could (assuming my relevant motivational strength did not mysteriously change after I had watched the news about the lottery). However, such a possibility does not immediately pose a threat to my account. In my view, it simply indicates that we may sometimes misunderstand our own hopes, wishes, and especially the order of priority among them. If my account is correct, the less reliable our answer to the focal question is, the more often we make a mistake in the corresponding self-ascription. This is an interesting empirically testable prediction drawn from my account.

So far, I have formulated three procedures for self-ascribing predominant, defeasible, and infeasible motivations, respectively. I have reached these procedures by describing, from our outward-looking perspective, how we think about what we are motivated for. The ways we think about it seem to subtly vary depending on the different modes of motivation. Hence, the resulting procedures also vary. The specific outward-looking descriptions I proposed regarding

the thoughts constituting those different modes of motivation may be controversial. However, I hope that they are at least enough to show that there is no fundamental difficulty in applying to desire the general schema of the successful transparency account specified at the beginning of this section. In my view, this is the most promising direction for transparency theorists of desire to pursue.

7. Conclusion

In this paper, I criticize the existing attempts to extend the transparency account of introspection to desire and suggest what I believe is a more promising direction to pursue. It is true that when we introspect our desire that *p*, we often find ourselves wondering whether *p* is desirable, appears desirable, is desirable in some specific respect, and so on. However, since answering these questions itself does not amount to desiring that *p*, it does not put us in a position to economically and authoritatively self-ascribe the desire. Instead, we should focus on the thought that the desire in question consists of. Describing this thought from the outward-looking perspective of the desiring subject leads to a method of desire self-ascriptions in which the ability of desiring is redeployed. Because of this redeploying nature, such a method, like that of Evans's, can explain the authority of the introspection of the desire at issue in a non-evidential and economical manner.

Acknowledgments

Earlier versions of this paper were presented at the 63th Annual Northwest Philosophy Conference in November of 2011, Nagoya Philosophy Forum 2012 Fall, the 64th Annual Northwest Philosophy Conference in October of 2012, and the 46th Annual Meeting of the Japan Association for the Philosophy of Science in November of 2013. I am grateful for helpful discussion with the audience. I am also thankful for the helpful comments from Robert Brandom, Jane Heal, Satoshi Kudo, Ulf Hlobil, Preston Stovall, Gurpreet Rattan, Takeshi Kanasugi, Michael Lockhart, Tuomo Tiisala, Kengo Miyazono, Masashi Kasaki, and Nebil Reyhani. This work was supported by a Grant-in-Aid for JSPS Fellows (10J07325) and the JSPS Overseas Research Fellowship.

References

Andreotta, A. J. (2020). Extending the Transparency Method beyond Belief: a Solution to the Generality Problem, *Acta Analytica*, online first.

Anscombe, G. E. M. (1963). *Intention*, 2nd edn. Cambridge: Harvard University Press.

Armstrong, D. (1993). *A Materialist Theory of the Mind*, 2nd edn. London: Routledge.

Ashwell, L. (2013). Deep, Dark…or Transparent? Knowing Our Desires, *Philosophical Studies*, 165, 245–256.

Baker, D. (2015). Why Transparency Undermines Economy, *Synthese*, 96, 3037–3050.

Bar-On, D. (2004). *Speaking My Mind*. Oxford: Oxford University Press.

Barz, W. (2015). Transparent Introspection of Wishes, *Philosophical Studies*, 172, 1993–2023.

Boyle, M. (2009). Two Kinds of Self-Knowledge, *Philosophy and Phenomenological Research*, 78, 133–164.

—(2011). Transparent Self-Knowledge, *Proceedings of the Aristotelian Society*, 85, 223–241.

Bratman, M. (1987). *Intention, Plans, and Practical Reason*. Cambridge, Mass: Harvard University Press.

Byrne, A. (2005). Introspection, *Philosophical Topics*, 33, 79–104.

—(2011a). Knowing What I Want. In J. Liu and J. Perry (Eds.), *Consciousness and the Self: New Essays* (pp. 165–183). Cambridge: Cambridge University Press.

—(2011b). Transparency, Belief, Intention. *Proceedings of the Aristotelian Society, Supplementary*, 85, 201–221.

—(2018). *Transparency and Self-Knowledge*. Oxford: Oxford University Press.

Carruthers, P. (2011). *The Opacity of Mind: An Integrative Theory of Self-Knowledge*. Oxford: Oxford University Press.

Castañeda, H.-N. (1974). *The Structure of Morality*. Springfield, IL: Charles C. Thomas.

Davidson, D. (1963). Action, Reasons, and Causes, *Journal of Philosophy*, 60. Reprinted in his *Essays on Actions and Events* (pp. 3–19). Oxford: Oxford University Press, 1980.

Edgley, R. (1969). *Reason in Theory and Practice*. London: Hutchinson.

Evans, G. (1982). *The Varieties of Reference*, Ed. John McDowell. New York: Oxford University Press.

Fernández, J. (2007). Desire and Self-Knowledge, *Australasian Journal of Philosophy*, 85, 517–536.

—(2013). *Transparent Minds: A Study of Self-Knowledge*. Oxford: Oxford University Press.

Finkelstein, D. (2003). *Expression and the Inner*. Cambridge, MA: Harvard University Press.

Gallois, A. (1996). *The World Without, the Mind Within*. Cambridge: Cambridge University Press.

Gertler, B. (2010). *Self-Knowledge*. London and New York: Routledge.

—(2011). Self-Knowledge and the Transparency of Belief. In A. Hatzimoysis (Ed.), *Self-Knowledge* (pp. 125–145). Oxford: Oxford University Press.

Goldman, A. (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, 1–14.

Gordon, R. (1996). 'Radical' Simulation. In P. Carruthers and P. K. Smith (Eds.), *Theories of Theories of Mind*. Cambridge: Cambridge University Press.

Grice, P. (1971). Intention and Uncertainty, *Proceedings of the British Academy*, 57, 263–279.

Hampshire, S. (1975). *Freedom of the Individual*, expanded edn. Princeton: Princeton University Press.

Lycan, W. G. (1996). *Consciousness and Experience*. Cambridge, MA: MIT Press.

Marks, J. (1986). Introduction: On the Need for Theory of Desire. In J. Marks (ed.) *The Ways of Desire* (pp. 1–15). Chicago: Precedent Publishing, Inc.

Moran, R. (2001). *Authority and Estrangement: An Essay on Self-Knowledge*. Princeton: Princeton University Press.

—(2003). Responses to O'Brien and Shoemaker, *European Journal of Philosophy*, 11, 402–419.

Nagel, T. (1970). *The Possibility of Altruism*, Princeton: Princeton University Press.

Nichols, S. and Stich, S. (2003). *Mindreading: An Integrated Account of Pretence, Self-Awareness, and Understanding Other Minds*. Oxford: Oxford University Press.

Paul, S. K. (2012). How We Know What We Intend, *Philosophical Studies*, 161, 327–346.

—(2014). The Transparency of Mind, *Philosophy Compass*, 9, 295–303.

—(2015). The Transparency of Intention, *Philosophical Studies*, 172, 1529–1548.

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

Samoilova, K. (2016). Transparency and Introspective Unification, *Synthese*, 193, 3368–3381.

Scanlon, T. M. 1998. *What We Owe to Each Other*, Cambridge: Harvard University Press.

Scarantino, A. and de Sousa, R. (2018). Emotion, In E. N. Zalta et al. (Eds.), *Stanford Encyclopedia of Philosophy*, retrieved from https://plato.stanford.edu/archives/win2018/entries/emotion/

Setiya, K. (2011). Knowledge of Intention. In A. Ford, J. Hornsby, & F. Stoutland (Eds.), *Essays on Anscombe's Intention* (pp. 170–197). Cambridge: Harvard University Press.

Shoemaker, S. (1996). Self-Knowledge and "Inner Sense." Lecture II: The Broad Perception Model. In his *The First-Person Perspective and Other Essays* (pp. 224–245). Cambridge: Cambridge University Press.

Schroeder, T. (2004). *Three Faces of Desire*, Oxford: Oxford University Press.

Schueler, G. F. (1995). *Desire: Its Role in Practical Reason and the Explanation of Action*. Cambridge, MA: MIT Press.

Shimamura, S. (2011). Knowing One's Own Mind: Resolving a Philosophical Dilemma concerning Self-Knowledge of Propositional Attitudes, Ph.D. Dissertation submitted to The University of Tokyo. (In Japanese).

Sinhababu, N. (2013). The Desire-Belief Account of Intention Explains Everything, *Nôus*, 47, 680–696.

Smith, M. (1994). *The Moral Problem*, Oxford: Blackwell Publishing.

Stocker, M. (1979). Desiring the Bad: An Essay in Moral Psychology, *Journal of Philosophy*, 76, 738–753.

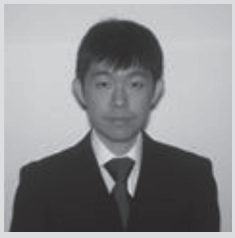
Valaris, M. (2014). Self-Knowledge and the Phenomenological Transparency of Belief, *Philosophers' Imprint*, 14, 1–17.

Velleman, D. (1992). The Guise of Good, *Nôus*, 26, 3–26.

Way, J. (2007). Self-Knowledge and the Limits of Transparency, *Analysis*, 67, 223–230.

Wright, C. (1998). Self-Knowledge: The Wittgensteinian Legacy. In C. Wright, B. C. Smith, & C. MacDonald (Eds.), *Knowing One's Own Mind* (pp. 13–46). Oxford: Clarendon Press.

About the Author



Shuhei Shimamura received his BA degree from Nagoya University, Japan, in 2004, his MA degree from the University of Tokyo, Japan, in 2006, and his Ph.D. degree from the University of Tokyo in 2011. Since 2017, he has been an assistant professor at Nihon University. His research interests include self-knowledge, semantic externalism, inferentialism, and logical expressivism.

✉ shuhei.shimamura0803@gmail.com