

Deep, dark...or transparent? Knowing our desires

Lauren Ashwell

Published online: 11 May 2012
© Springer Science+Business Media B.V. 2012

Abstract The idea that introspection is transparent—that we know our minds by looking out to the world, not inwards towards some mental item—seems quite appealing when we think about belief. It seems that we know our beliefs by attending to their content; I know that I believe there is a café nearby by thinking about the streets near me, and not by thinking directly about my mind. Such an account is thought to have several advantages—for example, it is thought to avoid the need to posit any extra mental faculties peculiar to introspection. In this paper I discuss recent attempts to extend this kind of outwards-looking account to our introspective knowledge of desire. According to these accounts, we know our desires by attending to what in the world we judge to be valuable. This, however, does not deal satisfactorily with cases where my value judgments and introspective knowledge of my desires come apart. I propose a better alternative for the proponent of transparency, but one that requires giving up on the supposed metaphysical advantages.

Keywords Self-knowledge · Desire · Transparency · Introspection

Despite the fact that the term “introspection” is used to name the distinctive method by which we know our own mental states, it is now widely held that we do not look *inwards* to find out what we believe. Instead, it is thought that we look *outwards* towards the world—towards the belief’s content. Self-knowledge comes from looking *away* from the self. Recent work has tried to extend this outward-looking account from belief to desire. In this paper I argue that these existing attempts fail; however, I provide a sketch of a more promising outward-looking account.

L. Ashwell (✉)
Department of Philosophy, Bates College, 2 Andrews Rd, Lewiston, ME, USA
e-mail: lashwell@bates.edu

I will begin by roughly outlining what it means for an introspective method to be *transparent*, focusing on introspection of beliefs. Following that, in Sect. 2, I look at the recent attempts to extend this idea to our introspective knowledge of desire, and I argue that none succeed in characterizing the method via which we come to have introspective beliefs about what we want. There is not, as of yet, a successful account of a transparent method for desire introspection. In Sect. 3, I explore the possibilities for giving a more plausible transparent account of desire introspection, drawing on insights gained from the failures of existing accounts.

1 Transparency

An outward looking method for self-knowledge seems plausible when we start with belief. Rather than turning our attention towards our minds in order to find out what we believe, we think about the content of the belief and whether that content is true. Gareth Evans describes the method as follows:

[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?” (Evans 1982, p. 225)¹

Transparent accounts of introspection reject the idea that we know our minds via an “inner-eye”—we don't focus on our mental states *qua* mental states in introspection. They also reject any other quasi-perceptual form of acquaintance with our mental states,² such as one might have if one had an internal scanning mechanism as posited by Armstrong (1968, Chap. 15), and, more recently, Nichols and Stich (2003, pp. 160–164).

However, not just any externally directed method will count as being transparent—it is also required that I do not attend to my behaviour. In order for me to know what *you* want or believe I must observe how you behave, and infer from that what your beliefs and desires, or other mental states, might be. Yet for me to know what *I* want or believe it seems that I don't *require* behavioural evidence.

¹ There are some extra complications with Evans's way of characterizing transparency that I will ignore for the purposes of this paper. For example, Evans's way of formulating the idea of a transparent method makes it seem as if I can, in applying this method, gather new evidence. Evans' method is presumably not meant to involve looking for new information about the likelihood of a third world war—looking, say, in today's news for reports of new conflicts when you have not read today's news yet. This isn't a method for knowing what my mental states are *now*. Also, my answers to the two questions will not always line up. My answer to “Will there be a third world war?” might be “I don't know”, and thus my answer to “Do you think there is going to be a third world war?” will be “No, I don't think that” (but not because I think there won't be).

² That said, Fernandez (2007) claims that his account of introspection is transparent, yet his metaphysics of introspection in fact involves something quite like an internal scanning mechanism. Part of his method for knowing our desires involves a direct causal link, which is not inferential, between a desire-like mental state and the belief that one has a desire. See Ashwell (2009, Chap. 3) for further explanation.

Of course, sometimes I do know what I want from looking at how I behave—perhaps I may know that I quite fancy someone by interpreting my nervous behaviour around him or her as attraction. But since I require attention directed on my *behaviour* to come to this belief about my desires, such an epistemological method will not count as transparent.

So, unlike how we know about others' mental states, a transparent method will not involve attending to your own behaviour. And unlike internal scanning mechanisms, or "inner-eye" accounts of self-knowledge, a transparent method will not involve attending to anything mental. These two parts of transparency are made clear in a statement of transparency from Richard Moran's *Authority and Estrangement*:

A statement of one's belief about X is said to obey the Transparency Condition when the statement is made by considerations of the facts about X itself, and not by either an "inward glance" or by observation of one's own behavior (Moran 2001, p. 101).

Thus we can see transparency as involving two parts:

- (1) *Transparency of the self*: In investigating my mental states, I do not attend to my mental states or behaviour, except insofar as these figure in the content of the mental state.
- (2) *Content-directedness*: In order to investigate whether I have a mental state with content p, I direct my attention (in thought) towards p being the case.

There are a few advantages that a transparent model of introspection is supposed to have. Firstly, it is supposedly metaphysically *economical*—it does not require positing any internal scanning mechanism, or any other special faculty via which we know our mental states, such as an "inner-eye", beyond what is needed for ordinary non-introspective knowledge.³ Take belief, for example: if a transparent method for knowing my beliefs simply involves looking to whether the content is true, then I do not need to use any special faculty beyond what is needed to come to the belief itself.⁴

Moreover, this transparent method for knowing my beliefs explains why each of us has a special sort of reliable access to our own beliefs, and not to others'. George can reliably infer from p that he believes that p, whereas Sam's inference from p to *George believes that p* would not be reliable. In order for this second inference to be even somewhat reasonable, Sam would have to have some reason beyond p itself for thinking that *George* would believe p. And even though p's being true does not raise the likelihood of it being the case that I believe p, *my* having the proposition p *as*

³ However, as will become clearer in Sect. 3, not all transparent accounts will have this advantage—in the end, I think that the most plausible transparent account of desire introspection will not have this feature.

⁴ And also the ability to infer *from* that belief to the belief that you have it (although some transparent accounts decline to call this inference, most agree that it is something *like* inference). The general ability to perform inferences is needed, of course, for non-introspective knowledge.

part of my apparent evidence entails that I already believe *p*, and thus ensures that the inference from *p* to I believe that *p*, made by me, is reliable.⁵

But belief, of course, is just one of our mental states. I will look at whether we can extend this idea to our knowledge of our desires, as has recently been suggested. I will presuppose that the distinctive method via which we know our own beliefs is transparent. Given this, a transparent model of how we introspectively know our desires can look very attractive. It would be a step towards an integrated (i.e. that we have introspective access to our different mental states in similar ways) and economical (we avoid positing special introspective faculties) account of introspection.

2 Transparency for desires

In Evans' original description of a transparent method, he stresses the *direction of attention*. My attention, in inquiring about my beliefs, is not on my beliefs *as* mental states, but on their content. The same might seem *prima facie* plausible for other mental states as well:

If asked whether I am happy or wishing that *p*, whether I prefer *x* to *y*, whether I am angry at or afraid of *z*, and so on, my attention would be directed at *p*, *x* and *y*, *z*, etc. (Bar-On 2004, p. 106).

...often my eyes are...“directed outward – upon the world.” I can investigate my preferences by attending to the *beer* and the *wine*... (Byrne 2005, p. 100)

If I am asked (by myself or others) whether I want *p* to be the case, my attention will be directed at *p* being the case. To address the question, I do not try to, so to speak, scan my own mind in search of a state that I can identify as the relevant desire. Rather, I concern myself with the outside world by focusing on the intentional object of the desire (Fernandez 2007, p. 524).

There must be more to the story, however, as we still need to know what we do once our attention is directed towards the intentional object of the desire. The method cannot be exactly the same as the method by which we have introspective access to our beliefs. Although, in looking at whether I desire *p*, I don't inquire whether *p* *is* the case, I seem to attend to *p* *being the case*—in particular, to what the world would be like were *p* to be the case. When I form a belief about my desires via introspection, my attention is *content-directed*.

However, the claim of transparency must go further than just the claim that our attention is content-directed; our attention needs to not just be content-directed, but also not at all directed on our own mental states or behaviour. Content-directedness, as it has to be understood to be part of an account of self-knowledge for desire, does

⁵ Thus, it in fact explains two things: the fact that we have very different sorts of procedure for knowing our own beliefs and the beliefs of others, and the fact that we have a high degree of justification for our beliefs about our own beliefs (since the inference is highly reliable). See Byrne (2005) for more detailed discussion of this.

not itself entail that we don't also have *some* attention on these other things. If our method involved thinking about *p* being the case, and then monitoring ourselves to see whether this made us happy, or observing whether this made us smile, then the method would not be transparent.⁶

As I mentioned before, insofar as there is a transparent method for knowing our desires, transparency cannot require us to attend simply to *whether p is the case*, as we do for belief. So what do we do, when we want to know our desires? What do we do once our attention is focused on the object of potential desire?

Suppose you would like to know whether you want to go to the beach. According to transparent accounts, if I want to know whether I want to go to the beach, I don't look inwards, but consider the beach (or the activity of going to the beach, or the state of affairs of my being at the beach). This doesn't need to involve literally looking at the beach (I cannot see it, sadly, from where I am), but it does involve attending to the beach as a non-mental item (though I do attend to it *in thought*).

What features of (going to) the beach will I attend to? A common thing we do when thinking about our options and what we want regarding them is to weigh the pros and cons of each option. This is a way of attending to the aspects of an option that make it a favourable or unfavourable thing to do. I will attend to features of the activity of going to the beach—I will attend to the likelihood of the beach being crowded, whether it is warm enough, and perhaps the cost of getting there. I will ask whether going to the beach is a *good option*—whether it is *desirable*. So, a potential transparent account of introspection for desires is that we know our desires, in a first-personal way, via our judgments of *value* or *desirability* concerning the object of the potential desire.

This is in fact, roughly, what is suggested in recent accounts given by Richard Moran (2001), and Alex Byrne (2005, 2011a). Though the details of each account are slightly different, they both give a central role to judgments of value or desirability. It is through attending to what we judge to be valuable that we find out, introspectively, what we desire.⁷

2.1 Moran and estranged desire

Moran puts forward the view that introspective knowledge proceeds by considering the *deliberative* question of which mental states one *ought* to have. This method is transparent since the question of which mental states one *ought* to have is addressed by attending to the world. I know whether I believe that *p* by asking whether I ought to believe that *p*, and I answer this question in turn by asking whether *p*. But this general method is not restricted to belief, Moran argues—it extends to other mental states too. For example, I find out, introspectively, *that I am afraid* by considering

⁶ Unless we also had a transparent method for introspecting that we are happy—in that case, the method of looking to see if *p* makes us happy *would* be transparent.

⁷ Fernandez (2007) also provides an account of desire introspection with some similarities to Byrne's and Moran's, although judgments of value play a much less central role. In what follows I will focus on Moran and Byrne's work, but I will also touch on the consequences this discussion has for Fernandez's picture.

whether my fear is *appropriate*, and in doing so I consider the world and whether there is anything in it to fear right now. And I find out my desires by considering what in the world is worth desiring—what is *desirable*, what is *valuable*.

This method, Moran claims, captures the truly first-personal nature of introspection; he takes it to be important that introspection, *by its very nature*, puts us in contact with our own, and only our own, mental states. Only I, from my point of view, can find out whether I *believe that p* by simply considering the question of *whether p*. Since it is at least in principle (or at least in science fiction) possible for *my* mental states to be directly causally hooked up to *someone else's* beliefs about them, a mere causal link between mental states and introspective beliefs is not sufficient for capturing the special first-personal character of introspection. For similar reasons, Moran does not think that introspection could resemble the way we interpret and theorize about other's mental states, where we take their behaviour as *evidence*—even if the evidence in our own case consisted in private thoughts, feelings, and dreams. Although we can have such knowledge, any beliefs we have about our mental states via a method that is not essentially first-personal will constitute *estrangement* from our own minds, as "...this information may as well be about some other person, or about the voices in her head" (Moran 2001, p. 93).

Moran does think that we can have such estrangement from our desires—any desire that I find out about in any way other than considering whether its object is valuable will be such a desire. I will be estranged from the unconscious desires that I find out about on the therapist's couch through interpretation of my behaviour or dreams. But I am also estranged from my weak-willed desire to watch more television—however I find out about it, it is not through thinking about the *value* of watching, or whether I *ought* to want to watch yet another episode of another generic crime drama.⁸ Other desires, such as those associated with bodily functions like hunger, will also be estranged from us, as they "...may be experienced by the person as feelings that simply come over him. They simply happen" (Moran 2001, p. 114).⁹ However we know about these, it is also not through addressing the deliberative question of whether we *ought* to desire food, water, sex, or whatever the feeling that comes over us is directed toward.

Note the differences between all these cases, however. The unconscious desire I find out about on the therapist's couch is just a theoretical posit to me—it does seem that I am at an epistemic step removed from the desire. But if I am estranged from my weak-willed desire or my hunger-driven desire for food, the estrangement is not epistemic. Although my weak-willed desire to watch more television, instead of working, is not known by considering whether I *ought* to have the desire to watch more television, or whether I *value* watching more television, it doesn't feel like the desire is someone else's. I take ownership of the desire in feeling guilty about so

⁸ Moran considers the akratic desire of a gambling addict to gamble, and concludes that this desire is one the addict is estranged from.

⁹ Fernandez also takes the view that desires based in bodily need are known not through judgments of value, but instead through what he calls *urges*: "...states wherein we experience the fact that we are not in some state as unpleasant" (Fernandez 2007, p. 521). Fernandez, however, differs from Moran in thinking that such knowledge would still be transparent; however, he has a unique understanding of what transparency entails. See Ashwell (2009, Chap. 3) for discussion of this.

desiring, and judging that I ought not to feel so pulled to procrastinate. If we are to get a full picture of desire introspection, our introspective access to such desires needs to be accommodated in the theory.

Moran admits that he does not intend to give a theory that encompasses all of our mental states; although some of our desires are known through considering the deliberative question of what desires to have, which in turn are answered by asking what is valuable, others (the estranged ones) are known in some other way. However, if we need to, for example, posit a mere causal connection between my weak-willed desire and my introspective beliefs about it, why not think that all desires are in fact known in this way?

So although Moran has isolated some important features that contribute to the first-personal nature of mental states, I think that he has over-stated his case in claiming that this constitutes the epistemic method of introspection for these mental states. Instead, the lesson we ought to take is that the deliberative question arises whenever we take a mental state as our own (although I suspect the relevant question in many cases will be one that is less prescriptive than what mental states one *ought* to have), even though we know of our mental states through another method. Sometimes the deliberative question about whether one ought to have a desire comes *after* detection of the desire, and sometimes the deliberative question is a method for *forming* a desire—but it is not, in general, the introspective method for *knowing about it*.

But even if Moran can answer these concerns, there is still another left unaddressed. My prior argument had to do with desires that we know about through first-personal introspective methods other than judgments of value. Moran considers these cases, and claims that these are desires that we are estranged from. But there is a second way that our desires and value judgments can come apart: it seems that sometimes, although I judge some activity valuable, I know I don't want to do it. I know I really *should* get out of bed when my alarm rings in the morning, but I just don't feel like it—I know that I just don't want to. I often judge that the healthy option on the menu is good for me—that it is valuable in the respect that it is healthy—yet know that I don't want it. And when I procrastinate, it isn't just that I know that I want to watch television without judging that it has any value, I also know that I don't want to work—I don't want to take the valuable option. Without giving some answer to *this* Moran can only have given, *at best*, a partial description of desire introspection. If all that was involved in introspecting desire was judging what is valuable, then whenever I judge that ordering salad is healthy and so valuable, I ought to conclude that I desire it.

2.2 Byrne and defeasible desire

To deal with this problem¹⁰ without giving up on the idea that desire introspection proceeds via inference from judgments of value, one needs to specify conditions

¹⁰ To deal with the other problem of cases where we judge we want something that we don't judge valuable, Byrne suggests that the sense of "valuable" here needn't be taken to be quite as strong as it sounds.

under which this inference is *defeasible*.¹¹ Alex Byrne (2011a) takes this strategy, and suggests a defeater based on the consideration of one kind of case where our desires and our value judgments come apart. However, I will argue that Byrne's suggestion for a potential defeater does not cover all cases it needs to, and furthermore requires foreknowledge of our desires in order to apply it. Thus, it is unsuitable to play a role in an account of how we know our desires.

Here is how the story is meant to go: I judge that going out and exercising is valuable. However, I do not conclude from this that I want to exercise. If we generally know our desires through our judgments of value, why do we fail to make the inference in these cases? Here, Byrne observes, we can see that I have a contrary intention—I *intend* to stay on the couch, despite my judgment that exercising is valuable. But this can't be the whole story, as I can have contrary intentions in cases where I still know I have desires for the other option—suppose I want to order the ravioli yet I also have a desire for steak, and I form an intention to go with the pasta. This doesn't stop me from judging that I want steak—I just don't want it as much as I want ravioli. So we need further conditions beyond just that I have a contrary intention.

Byrne thinks that what is special about the couch vs. exercise case is that I judge the lazy option I intend to take—that of staying on the couch—to be *not* desirable. We add this together with the intention to stay on the couch to get the following defeater for the inference from the value of an option phi-ing to the belief that I desire to phi:

Desire Defeater: If you intend to psi (where phi-ing and psi-ing are believed incompatible) and you judge that psi-ing is not desirable, believe that you don't desire to phi.

So in the case where I believe that it is desirable to get up off the couch and get some exercise—yet I know I don't want to—I intend to stay lying on the couch, though I judge that undesirable. According to *Desire Defeater*, because I recognize that I intend to take the undesirable option of remaining on the couch, I will judge that I don't desire to exercise. The application of *Desire Defeater* explains why I don't believe that I want to take the option judged valuable.

Note that although this defeater appeals to *mental evidence*, it may still fit with our understanding of what it is for an introspective method to be transparent. If the method via which we know our intentions is transparent,¹² this defeater can still be part of a transparent method via which we introspectively know our desires—the actual defeater used will in fact appeal to the transparent method for knowing what we intend instead of directly appealing to judgments about intentions.

¹¹ Here I am interested in defeaters in the sense that they stop you from believing that you have the desire in question. One might also be interested in defeaters as mere justification-removers, if the particular interest was in introspective *knowledge*, rather than just introspection.

¹² See Setiya (2011) or Byrne (2011b) for a potential transparent method for knowledge of intention: roughly, I know what I intend via knowing in a non-observational way what I am doing. A discussion of whether this account is viable is beyond the scope of this paper, and I will here just assume that it is.

The problem is that I may not judge that staying on the couch is *undesirable*—so *Desire Defeater* will not cover all cases we need to account for.¹³ In some situations, it may be that staying on the couch has something going for it. It may even have quite a lot going for it. The couch is comfortable, and staying there takes less effort than getting up—but the couch’s having something going for it doesn’t make me judge that I want to exercise. I may still not want to exercise, despite judging it desirable. In this case, *Desire Defeater* does not apply because the option I intend to take (staying on the comfortable couch) is judged desirable. And not all problem cases need to be ones of *laziness*. I judge having children, in what would be a supportive, loving environment, to be valuable. However, I do not find myself with any desire to. Yet I don’t think that the childless path that I intend to take is *undesirable*. It has a lot going for it. There is a choice between two desirable options, and while one moves me, the other, currently, does not.

It may be that the account is simply incomplete—there may be other defeaters that Byrne has yet to identify. And defeaters may have their own defeaters, which account for cases where applying them would yield results that conflict with cases of introspection. One might wonder, however, whether Byrne’s defeater ever gets the right answer. After all, don’t we know our intentions *through* knowing what we want? I take *how much* I want something, or *whether* I want it, into account when deciding what to do. I need not form an intention to stay on the couch *in order* to know that I don’t want to exercise—the fact *that I don’t want to exercise* may be part of what I consider when deciding *whether* to intend to exercise or to stay on the couch. “Shall I go for a run?” I think, “...hmmm...it would be best, but *I just don’t want to*...so I’m not going to go!” If we introspectively know our desires *before* we know what we intend, then a defeater that appeals to knowledge of intention cannot be part of the story of how we have introspective access to our desires.

This objection, however, goes too far if it claims that desires are *always* epistemically prior to forming intentions. It seems clear that practical reasoning need not always involve overtly attributing desires to oneself, and in many cases one may just think about the value or desirability of options, the pros and cons, rather than what one wants. So we can agree with Byrne that desires rarely figure in our conscious consideration of some option. We usually think about the features of the option itself—we normally think about the value or desirability of options rather than our desires in forming intentions. However, this is only in the *normal* cases—the cases for which Byrne’s defeater does not apply, as they are the ones in which our introspective access to desires and our value judgments line up. When there is nothing to be said for what I intend to do and I could *instead* be doing something that I judge to be worthwhile, the only reason I can give for my intention is that *I just don’t want* to do the worthwhile thing. If there really is nothing to be said for staying on the couch, and going for a run is desirable, I will have to appeal directly to my desires in forming my intention to stay put. Thus I must know what I want before I know what I intend, and thus Byrne’s defeater cannot be part of the method via which I know what I want.

¹³ It may also, in fact, not account for *any* such cases. If Byrne weakens his sense of *judging valuable* enough that it can deal with cases where it may seem natural to say that I introspect that I desire something without judging it to be valuable (to, perhaps, *there is something to be said for it*), then it may be that anything we’d *intend* to do would be judged valuable in this weaker sense.

3 Transparency in a different way?

Neither Moran nor Byrne have a plausible answer to the problem of how we can fail to judge that we desire something that we nevertheless judge to be valuable. Is there any hope for a transparent account of desire? Here I will explore what one might look like, drawing from some of the lessons we learned from the failures of these existing transparent accounts. Instead of focusing on *judgments of value*, I suggest transparency theorists ought to be looking at something closer to *appearances of value*.

To explain what I mean, it will be helpful to think about an analogy with the connection between perceptual experience and our judgments about how things are near to us. Sometimes we judge that things are a certain way near us—that there is, say, a tree in front of us—even though we do not have perceptual evidence for this (the tree may be behind a fence). Such a case would be one where we believe this on the basis of testimony. Thus we judge that there is a tree in front of us even though this judgment is not made on the basis of perceptual appearances. Other times our perceptual experience and judgments of how things are around here diverge in a different way—we may judge that things aren't the way that experience tells us, such as when we have reason to believe that our perceptual experience is illusory.

Value judgments and introspective desire judgments come apart in a similar way to how judgments about *how things are around here* and judgments about the content of our perceptual experience may come apart. One way in which we come to believe something is desirable or valuable is via testimony. You might tell me that the sushi at this restaurant is good; my dentist might tell me that flossing is desirable. But though I may come to have appropriate beliefs about the value of flossing, I may, unfortunately, fail to desire to floss.

A difference between this direction of the connection in perception and the connection in desiring is that when I come to believe that things are a certain way around here, I don't *usually* come to experience them to be that way—whereas when I come to believe that things are valuable on the basis of testimony, often I *also* then come to desire them. But not always. And when I don't desire them, this is not necessarily a fact that is introspectively hidden from me. What I suggest is that in these cases, the thing does not *appear* to me to be valuable, although I judge it to be so.

Other times you judge something to be valuable, but know that you don't desire it, because although you *once* saw the thing as desirable and have retained a memory of that (and judge that the value of the thing has not changed), the thing now *no longer* appears to you to be desirable. I might believe that continuing my once-favourite hobby is valuable, although for the moment it seems to have lost its lustre. This case, in the analogy, is like continuing to believe that there is a tree in front of you even though you have now shut your eyes.

Moreover, I suggest, we can have things appear valuable or desirable to us, although we do not end up judging them to be so. There can be *known illusions of value*. In these cases we have competing evidence that such things are *not* desirable, just as the competing evidence of a ruler can stop you from judging that the Müller–Lyer lines are of different lengths although they still appear to be so. Another glass

of wine might look attractive to me, but I may believe that having it would not be good at all because I have evidence that having more would be a bad idea—and so, although I know that I want it (because it appears desirable), this is not because I judge it valuable.

So how does this deal with Byrne's couch potato case? It certainly explains why I fail to form introspective judgment that I desire to exercise—although, in this sort of situation, I judge that exercising is valuable, this is *not* through current value appearances. The rest of the story depends on how we flesh out the details of the case. In Byrne's original case, where lying on the sofa is judged to *not* be valuable, there was also a corresponding lack of introspective attribution of desire to stay on the couch.¹⁴ On the account put forward here, the explanation for failing to form an introspective belief that you desire to stay on the couch is simply that staying on the couch does not appear to be valuable. And this seems plausible—when I don't want to exercise, yet this is not because staying where I am is something that I judge I want to do, the option of staying where I am appears not to be appealing either.

If, on the other hand, it was the case that lying on the couch *appears* to have something going for it, then I instead judge that I want to stay on the couch.¹⁵ Thus one of the advantages of this account over one based on value judgments is that it can accommodate *both* the case where the option judged less valuable is (introspectively believed to be) desired, and *also* the one where it is not. A *judgment* that something is not valuable is compatible with both the appearance and non-appearance of value, just as it is also compatible with a desire or lack of desire for the thing judged to not be valuable.

One problem with developing an account of transparent desire introspection based on value appearances is that it seems to give up on metaphysical economy in two senses. First, we have to posit some faculty that allows us to see value, or at least *seem to see* value, in things. Second, I suspect that if this account is fleshed out, the most plausible way to develop the metaphysics will involve accepting that there is a causal connection from your desires to value appearances.¹⁶ Wanting things makes you see them in a certain light, and this is how you introspectively know what you want.

¹⁴ I assume that this is a feature of the original example since Byrne's account does not allow for introspectively judging that I want to stay on the couch despite judging it not to be valuable at all. If I do conclude that I want to stay on the couch, this must be through non-introspective means—perhaps by inference from behaviour (I'm still lying here, so I must want to stay on the couch).

¹⁵ This happens a lot for me with early morning exercise—I wake up and judge exercising valuable, but it just doesn't appeal—but staying in bed does just *seem* to be valuable, even though I judge that it is not.

¹⁶ If we do not have such a connection between desires and value appearances, we give up the supposed reliability of desire introspection, which is another thing that transparency theorists generally want to maintain. This connection, it seems to me, would be one that is peculiar to introspection. However, if desire (in its normal operation), involves a faculty for knowing about value, and so *already* requires a connection to value appearances, this account *could* be economical in the sense that the connection posited is required to explain non-introspective workings of the mind. See Setiya (2011) for a structurally similar claim about introspective knowledge of intention, which involves intention-formation, by its normal operation, being a faculty for knowing what one is doing.

4 Conclusion

To be transparent, a method for knowing our mental states must be, at least, outward looking and ignore the self (except insofar as the self figures in the content of the mental state). If we had introspective access to our desires via *judgments* of value or desirability, we would have a transparent account of desire introspection. However, our value judgments and our introspective desire judgments very often do not line up. Moran gives us no account of how we know that we don't desire something that we value, and thus fails to describe the introspective method via which we know what we want. And although Byrne does attempt to account for these cases, the defeater he proposes does not work. Byrne's defeater does not cover all cases of desire-free valuing—but ultimately the problem with the defeater is that it is via knowing our desires that we know we are in the defeater's antecedent conditions, thus making this defeater unsuitable for playing a part in an account of how we have first-personal access to our desires. Those who want to give a transparent method for knowing our desires should focus instead on *appearances* of value rather than judgments—however, doing so will probably require giving up on metaphysical economy, which may remove their motivations for believing in transparent introspection. A commitment to value *appearances* might seem too much to swallow. If so, however, I think we ought to give up on a commitment to transparency.

Acknowledgments Many thanks to Alex Byrne, Richard Holton, Rae Langton, Charlotte Witt, Mark Okrent, Willem deVries, Susan Stark, Sarah Conly, and also audiences at the Massachusetts Institute of Technology, the University of New Hampshire, and the 2010 Northwest Philosophy Conference at Willamette University.

References

- Armstrong, D. (1968). *A materialist theory of the mind*. London: Routledge and Kegan Paul.
- Ashwell, L. (2009). *Desires and dispositions*. Ph.D. thesis, MIT, Cambridge, MA.
- Bar-On, D. (2004). *Speaking my mind: Expression and self-knowledge*. Oxford: Oxford University Press.
- Byrne, A. (2005). Introspection. *Philosophical Topics*, 33(1), 79–104.
- Byrne, A. (2011a). Knowing what I want. In J. Liu & J. Perry (Eds.), *Consciousness and the self: New essays*. Cambridge: Cambridge University Press.
- Byrne, A. (2011b). Transparency, belief, intention. *Proceedings of the Aristotelian Society, Supplementary Volume 85*(1), 201–221.
- Evans, G. (1982). *Varieties of reference*. Oxford: Oxford University Press.
- Fernandez, J. (2007). Desire and self-knowledge. *The Australasian Journal of Philosophy*, 85(4), 517–536.
- Moran, R. (2001). *Authority and estrangement*. Princeton, NJ: Princeton University Press.
- Nichols, S., & Stich, S. (2003). *Mindreading: An integrated account of pretence, self-awareness, and understanding other minds*. Oxford: Oxford University Press.
- Setiya, K. (2011). Knowledge of intention. In A. Ford, J. Hornsby, & F. Stoutland (Eds.), *Essays on Anscombe's intention*. Cambridge, MA: Harvard University Press.