

Knowing Our Desires

Despite the fact that the term “introspection” is used to name the distinctive first-personal method via 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 – we attend to the objects and properties and states of affairs that our beliefs are about. If I want to know, for example, whether I believe that there is coffee in the cupboard, I think about the cupboard and its contents, not about my own mind. The rough idea here is that we *see through* our mental states to the world; our investigation of our mental states is *transparent* to the world. Self-knowledge, then, 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 introspective knowledge of our desires often requires us to pay attention to ourselves, and not just the external world.

An outward looking method for self-knowledge seems plausible when we start with belief. To find out what we believe, we think about the content of the belief and whether that content is true. As expressed by Evans:

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: 225) ¹

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. But this isn’t a method for knowing what my mental states are *now*. See Gertler, 2007 for this sort of criticism.

This kind of outward looking account is now generally known as a *transparent account*. Unlike how we know about others' mental states, a transparent method will not involve attending to your behavior. And unlike internal scanning mechanism 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: 101).

Transparent Accounts of Desire Introspection

Although Evans and Moran here state the Transparency Condition in a way that explicitly restricts it to belief, this condition has been thought to extend to 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: 106).

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: 524).

...[in introspecting my preferences] often my eyes are still "directed outward – upon the world." I can investigate my preferences by attending to the *beer* and the

wine... (Byrne 2005: 100)

There's still more to say here – what exactly do we do once we have fixed our attention on the beer and wine? Well, one thing we sometimes seem to do when we think about what we want is to look at the pros and cons of an option – things which make it a good thing to do, or make it valuable, or desirable (or undesirable). And this is in fact what has recently been suggested as a transparent account of introspective knowledge of our desires, by Alex Byrne and Jordi Fernandez.² Very roughly, their accounts claim that we know our desires through inference from judgments of value.³

However, any account that seeks to explain our knowledge of our desires via an appeal to judgments about value will have two *prima facie* problem cases to face:

- 1) **Desire-free valuing:** cases where you judge something valuable, but know you don't want it. For example, I might judge it valuable to exercise, yet know in a first-personal way that I don't want to.
- 2) **Valuing-free desiring:** cases where you know you desire something but you don't judge it to be valuable. For example, I might know that I desire to have another drink, but I think it's a bad idea – I do not judge it valuable.

One way to deal with Valuing-free desires is to posit a further way in which we epistemically access our desires; this is the route taken by Fernandez. On the other hand

² See Fernandez, 2007; Byrne, 2005: 99-100.

³ This characterization is more correct as one of Byrne's view than Fernandez's. Fernandez does not think that inference is involved at all (see Fernandez 2003 for an argument to this effect), but instead that value judgments simply *cause* the belief that you have the relevant desire. This, however, involves positing a psychological mechanism (the brute causal connection between the valuing and the introspective belief) that is peculiar to introspection – something Fernandez claims he avoids. This difference does not affect the arguments in this paper.

you might deny that such desires exist, as Byrne does, and point out that the sense of “valuing” here need not entail thinking the thing in question to be very important; we can find having a cup of tea, or singing in the shower, valuable in the sense that there is *something to be said* for doing it.

This still leaves the problem of Desire-free valuing – very often we judge that something is valuable yet know in a first personal way that we don’t desire it, which becomes even more urgent once we weaken the sense of “valuing” to overcome the problem of valuing-free desiring – there are many things that I may judge have something to be said for them, but I still know that I don’t desire them. Burning precious family photos might keep me warm on a cold winter’s night, but I know that I have no desire to do so.

Neither Byrne nor Fernandez adequately addresses either of these problems. But I think that these accounts of introspection for desires based on judgments of value fail in instructive ways. If we focus on the cases where value judgments and desire judgments come apart, we can begin to build a more promising transparent account of desire introspection, which I will outline next.

An Alternative: The Projective Account

When I am lying on the couch, judging that exercise is valuable yet knowing I don’t want to, exercise just doesn’t *seem attractive to me*. I judge that it’s valuable – I have beliefs to this effect – my doctor, my friends, the New York Times all tell me that it’s valuable, but I just don’t feel it. On the other hand, when I desire to do something that I fail to judge to be valuable – say, when I know that I want another drink but judge that it would be a bad idea

– the drink *appears* to me as desirable or valuable – I just have other reasons to think that this appearance is not to be trusted – ignoring the reasons to the contrary, it might *seem* like a good idea, but I have defeating evidence – thus I judge that it’s not the thing to do. I think that an account based on *appearances* of value – rather than judgments – is a more promising account of desire introspection.

The idea behind this account is, roughly, that when looking out to the world (in thought or in cases of actually looking) *things appear to us to be desirable*. In the same way that the chocolate cake looks brown, it also looks *tasty* (thus desirable in the tastiness sense). Yet this account is *transparent* because our attention is on the object of desire, and not on ourselves.

The problem with taking inference from judgments of desirability or value as the method via which we know our desires stems from the fact that there are several *different* ways in which we can end up with a belief that something is desirable – and it is generally the way in which we come to this belief, and how that belief is sustained, that affects the reliability of the inference from value to desire. 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. When I come to believe that things are valuable on the basis of testimony, sometimes I also then come to desire them. But not always. This is what I think happens in some cases of desire-free valuing – I come to believe that something is valuable but this fails to generate in me the appropriate desire.

On this view, there are two kinds of desire-free valuing: firstly, there are cases where you have come to believe that something is desirable through outside sources, although

you do not see the thing *as* desirable – such as when my dentist tells me it is desirable to floss. Secondly, there are cases where 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), although the thing now *no longer* appears to you to be desirable. For example, I might believe that continuing my once-favourite hobby is valuable, although for the moment it seems to have lost its lustre.

Valuing-free desiring, on the other hand, can be accounted for as cases where we see something as desirable yet do not judge it to be so, perhaps because we have competing evidence that such things are *not* desirable. Another glass of wine might *look* attractive to me, but I may believe that having it would not be valuable 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.

On the projective account, it is in seeing the thing *as desirable* that I have access to whether I desire it; it is on the basis of this sort of evidence that I form beliefs about my desires, not on the basis of my desirability *judgments*. Only in a certain subset of cases where I judge some option desirable will I believe that I desire it – those in which the object of desire also *appears* to me to be desirable.

4.1 Evaluating the Projective Account

However, although the projective account fares better at accommodating the cases where desires and value judgments come apart, it still does not completely succeed as a transparent method for introspecting our desires. Firstly, there is a difference between wanting something and merely liking it. People may show all the signs of liking

something, yet have no motivation to get it – and *vice versa*. Appearances of desirability or value seem to fit better with judgments of what one *likes* than what one feels motivation to get. If motivation and liking are realized by different physical systems, as is suggested by Berridge (2004) and Robinson and Berridge (1998), then it is possible to not feel motivated to do the very thing that appears to you to be valuable. This is a threat to the projective account if, as seems likely, people with decoupled liking and motivational systems would *know* (first-personally) that they lack motivation for things they like.

Even if the properties projected were something like *attraction* – something like a magnetic *pull* – to have access to whether you feel pulled towards you would need to direct your attention inwards. To know what one is *motivated* to get or do one has to look inwards, to an inner push or pull.

This account also faces a challenge from experiments on early childhood theory of mind. If we inferred what we want from appearances of desirability, such that those appearances seemed to be desire-independent objective properties of the desire's object, we would expect children to often make mistakes and attribute their own desires to others. Yet at roughly the same age that they start talking about desire, they recognize that different people want different things.

This is in contrast to their talk of belief, where they often mistakenly attribute their own beliefs to others – young children have been found to not attribute beliefs to others that they know to be false, even if it seems that they have very good reason to think that the other person would have a false belief. Contrasting their own beliefs with those of others typically does not come until after age 3. Yet they are quicker to realize that different

people want different things,⁴ which suggests that they don't come to beliefs about their desires via appearances of desirability that seem to be objective properties of the desired object.⁵

In the end, I think, we can't have an account of desire introspection that has us only looking outwards to the world. Instead, we need to have some attention on ourselves. Although it is the most promising transparent account of introspection, the projective account has several problems. It does not fit with research on desire attribution by young children. It also obscures the difference between wanting and liking, as projective appearances of desirability match up best with what we like rather than what we feel motivated to do. Felt motivation, an inner push or pull that we experience in looking for what we want, requires us to look at ourselves, not just at the world. Though we do have to have our attention on the object of desire, in order to know whether we want it, we also need to keep one eye on ourselves.

References

- Bar-On, Dorit. 2005. *Speaking My Mind: Expression and Self-Knowledge*. Oxford University Press.
- Bartsch, Karen and Henry Wellman. 1995. *Children Talk About the Mind*. New York: Oxford University Press.

⁴ See (Bartsch and Wellman 1995). Subjective contrastives (talk that shows that the child recognizes that a mental state is subjective) for desire precede subjective contrastives for belief (Bartsch and Wellman 1995: 100). They also say that young children understand that "...desires are experienced by persons, and persons have differing desires. In our data young children certainly do not presume that desires are shared by all and sundry, nor do they egocentrically think only of their own desires without recognition of the conflicting desires of others. To the contrary, discussing recognized conflicts over desires is common in the speech of quite young children" (Bartsch and Wellman 1995: 83-85). This comes at around age 2 1/2.

⁵ "An objective misconstrual would view desirability as inherent in the object, not subjects; the intentional object of mental states are simply desirable (or not) in themselves, and hence experienced similarly by all who encounter them" (Bartsch and Wellman 1995: 85). They do not see this confusion in young children's talk about desires.

- Berridge, Kent. 2004. "Pleasure, Unfelt Affect and Irrational Desire". In *Feelings and Emotions: The Amsterdam Symposium*, A. Manstead, N. Frijda and A. Fischer (ed.), 243–62. Cambridge: Cambridge University Press.
- Byrne, Alex. 2005. "Introspection". *Philosophical Topics* 33: 79-104.
- Evans, Gareth. 1982. *Varieties of Reference*. Oxford University Press.
- Fernandez, Jordi. 2003. "Privileged Access Naturalized". *The Philosophical Quarterly* 53: 352-72.
- Fernandez, Jordi. 2007. "Desire and Self-Knowledge". *The Australasian Journal of Philosophy* 85: 517-36.
- Gertler, Brie. 2007. "Do We Determine what We Believe by Looking Outward?"
Available at SSRN: <http://ssrn.com/abstract=996267>
- Moran, Richard. 2001. *Authority and Estrangement*. Princeton University Press.