teorema
Vol. XXXIV/1, 2015, pp. 120-134
ISSN: 0210-1602
[BIBLID 0210-1602 (2015) 34:1; pp. 121-134]

Fernández on Transparency

André Gallois

Jordi Fernández's *Transparent Minds* is an important contribution to a debate that has attracted a great deal of attention over the last three decades. The debate focuses on two questions: what kind of epistemic access do we have to our consciously held psychological states, and what enables us to have such access? One answer to these questions is given by the introspectionist. According to the introspectionist we come to know what desires and beliefs we have through exercising a kind of inwardly directed perception. Fernández's book is one of the best defenses of a different answer: sometimes referred to as a transparency account of self-knowledge. It is ingeniously argued. Despite that I have two main reservations about it. The book is divided into two parts. In the first, what Fernández calls the Bypass view is developed and defended. In the second, Bypass is applied to solve three philosophical problems. My first reservation is that Bypass rests on an inadequate conception of justification. My second is that Bypass' role in solving the philosophical problems to which it is applied seems to be redundant to solving them.

I

What has triggered the debate, which is the topic of Fernández's book, is that our knowledge, or justified belief, about our own propositional attitudes has two distinctive features. As Fernández, following Alex Byrne, puts it, it is both special and strong. Special in that it does not have the same basis as the knowledge, or justified belief, that others have about us. Strong in that we, typically, have greater authority than others about what consciously held psychological states we are in. Fernández sums this up by saying that our epistemic relation to our

consciously held psychological states is privileged. Moreover, when we examine the basis for self-attributing beliefs, desires, and other propositional attitudes, specialty seems to conflict with strength. They appear to conflict because our self-attributions of propositional attitudes seem to be special in that they are based on nothing. If they are, far from being privileged, such self-attributions should, on the face of it, be given little or no epistemic weight.

Bypass is designed to show why self-attributions of propositional attitudes are privileged. It does so by attempting to show that they do have a basis. Their basis is the very same as the basis for holding the propositional attitudes that are self-attributed. On the basis of my present sensory intake I am justified in believing that there is a blue coffee cup in front of me. According to Bypass the same sensory intake justifies me in believing that I believe there is a blue coffee cup in front of me. More generally, reasons for believing that P double up as reasons for self-attributing the belief that P.

When applied to self-attributions of belief, Fernández' main argument for Bypass rests on the following conception of adequate support. He says:

> On the notion of epistemic justification that I will be using, a subject's state qualifies as adequate support for one of her beliefs when the state is of a kind that, in that subject, tends to correlate with the type of state of affairs that makes the belief true' [Fernandez (2013), p. 43].

My blue coffee cup-like sensory intake adequately supports my belief that there is a blue coffee cup in front of me since that type of sensory intake correlates with there being a blue coffee cup in front of one.

With this conception of adequate support in place, Fernández has no trouble showing that if the belief that P adequately supports the belief that Q, then the belief that P adequately supports the belief that one believes that Q. All he needs is that there be a high correlation, which surely there is, between consciously believing something, and, at least on reflection, believing that one believes it. If that is so, then there will be a high correlation between self-attributing a belief and whatever makes the self-attributed belief true. A least that will be so in a world like ours. It will not be so for a brain in a vat, which makes one wonder whether, given Bypass, a brain in a vat can have justified beliefs about what it believes.

One problem with this argument for Bypass is that it relies on a view of epistemic support that is, at best, highly controversial. Essentially it relies on a straightforward regularity account of justification that has been subjected to numberless counterexamples over the years. Setting that aside there seem to me compelling reasons for denying that what justifies self-attributing a belief is the very justification one has for holding the self-attributed belief. Here is one. Suppose Jones wants to see a certain film. Jones believes that it will be showing at her local cinema because the cinema has announced that it will. Further, Jones believes that her only justification for believing that the film will be showing is the cinema's announcement. Now, suppose a friend maintains that the cinema made no such announcement. Imagine Jones responds to her friend with the following:

> My sole justification for believing that the film will be showing is that the cinema announced that it will be. But, the film will be showing even if the cinema did not announce that it will be.

Something has gone badly wrong. What is it? In making her announcement Jones is incoherently endorsing the following pair of claims:

> (a) My sole justification for believing that the film will be showing is that the cinema announced that it will be,

and:

> (b) But the film will be showing even if the cinema did not announce that it will be.

If Jones is rational, and has an adequate grasp of the concept of justification, he will concede that one of (a) or (b) has to be rejected. But which one is an open question. Jones may concede that (a) is false. There is an independent justification for believing that the film will be showing, or he is simply not justified in believing that it will. Alternatively, Jones may continue to endorse (a), but reject (b). The crucial point is that he cannot have it both ways. If he continues to endorse both (a) and (b), he is allowing that the truth of his supposed justification for believing that the film will be showing is irrelevant to its actually justifying that belief.

Now suppose that Jones makes a second pronouncement. He says:

> My sole justification for believing that I believe that the film will be showing is that the cinema announced that it will be. But, I believe the film will be showing even if the cinema did not announce that it will be.

Again Jones is incoherently endorsing a pair of claims. This time they concern the justification he has for self-attributing the belief that was the topic of his first announcement. They are:

> (a') My sole justification for believing that I believe that the film will be showing is that the cinema announced that it will be,

and:

> (b') But, I believe the film will be showing even if the cinema did not announce that it will be.

In this case Jones is confronted with a like choice. He cannot hold on to both (a') and (b'). So he has to choose which to give up. In this case there is no question which one he should reject. It is (a') rather than (b'). But in rejecting (a') he is allowing that either the cinema's announcement provides no justification for his self-attribution, or else that there is an independent justification for making it.

What does this show? It does not show that the justification for holding a first-order belief cannot justify holding the relevant second-order one. What it does show is that if Bypass supplies a justification for making self-attributions of belief, there has to be a way to justify making such self-attributions that is independent of Bypass. If that is so, we should look for a more fundamental way of accounting for the relevant self-knowledge than is provided by Bypass

Is there an argument to show that Bypass does not yield any justification for making self-attributions of belief even one that is not fundamental? I think there is. It is this. Suppose it is stated in a highly reputable historical text that Constantinople fell in 1453. That it does so justifies Samantha in believing Constantinople fell in 1453. Since that is so, she is justified in believing the following conditional: If it says in the historical text that Constantinople fell in 1453, then Constantinople fell in 1453. In general, if P justifies someone in believing that Q, then she is justified in believing: If P then Q. According to Bypass, that it says so in the historical text justifies Samantha in believ-

ing that she believes Constantinople fell in 1453. If so, Samantha should be justified in believing: If it says in the historical text that Constantinople fell in 1453, then she (Samantha) believes Constantinople fell in 1453. But Samantha is not justified in believing that conditional.

Call the account of justification that Fernández is employing to defend Bypass the Bypass Account. What the previous pair of arguments indicate is that the Bypass Account is too undemanding. Something needs to be added to it. So, what needs to be added to the Bypass Account? Suppose, we think of reasons for belief as beliefs. It is not enough for the belief that P to justify, or serve as the basis for, the belief that Q that there be a high correlation between Q and believing that P. In addition, there needs to be a suitable connection between the content of the belief that P and the belief that Q. P must increase the likelihood of Q. That is why Constantinople falling in 1453 does not justify Samantha believing that she believes that it fell then. Constantinople falling in 1453 does not increase the likelihood of Samantha believing that it fell then.

Of course, Constantinople falling in 1453 might justify Samantha believing that she believes that it did. Suppose Samantha is a Byzantine historian. Put together Samantha's belief that she is a Byzantine historian with her belief that Constantinople fell in 1453 and you obtain a justification for her believing that she believes Constantinople fell in 1453, As she would put it, the justification would go something like this. If Constantinople fell in 1453, it would say so in histories of the Byzantine Empire. As a Byzantine historian I would have read histories of the Byzantine Empire. So, if it says that Constantinople fell in 1453 in histories of the Byzantine Empire, I would have read that it did, and presumably believed it.

Call this the Byzantine historian's justification. Note that a justification of this sort cannot be what Fernández has in mind when he would claim that the belief that Constantinople fell in 1453 justifies someone who believes that it did in self-attributing that belief. The justification for the self-attribution of a consciously held belief is supposed to be special in that it is a justification that no one else can employ. It is also supposed to be strong in that it normally overrides the justification that anyone else has to deny that self-attribution. But, there is nothing special or strong about the Byzantine historian's justification. Exchange the first person indexical for a proper name of the historian and anyone can employ it to attribute the relevant belief to

the historian. Moreover, there is nothing overriding about the Byzantine historian's justification whether or not it is employed to make a self-attribution or an attribution to another.

Fernández extends the application of Bypass to self-attributing desires. Basically, the same strategy is employed as with belief. I want an ice cream. I believe that if I go to the local shop, I will obtain an ice cream. So, I want to go to the local shop. My justification for wanting to go to the local shop is the combination of the desire for an ice cream with the belief that I will get an ice cream there. According to Bypass, the same combination of desire and belief justifies me in believing that I want to go to the local shop. It does so because of the high correlation between desiring to go to the local shop and believing one has that desire. Put that correlation together with the correlation between the previously mentioned desire- belief pair and desiring to go to the shop, and you get a high correlation between the desire-belief pair and believing one desires to go to the local shop.

The Bypass account of justifiably self-attributing desires rests on the conception of justification that was criticized previously. I am faced with the following choice. Either reject:

> (a*) My sole justification for believing that I want to go to the local shop is that I want an ice cream and if I go to the local shop, I will get one,

or:

> (b*) Even if it is false that I want an ice cream, and will get one if I go to the local shop, I still want to go to the local shop.

Again the choice between rejecting (a*) or rejecting (b*) is, as they say, a no brainer. It is (a*) that should go. Whether or not going to the local shop will result in an ice cream, given my possibly false belief that it will, I do want to go to the local shop. So, I should reject (a*). But rejecting (a*) is bad news for Bypass as a fundamental account of the justifiable self-attribution of desires. If I am justified in believing that I want to go to the local shop, there must be some justification for holding that belief which is independent of Bypass.

II

Before giving his account of self-knowledge, Fernández sets out a number of desiderata that any such account should, in his view, satisfy. Prominent among them are the following. It must explain [Fernández (2013), p. 38]:

(i) Why we have special access to our mental states when we self-attribute them,

and:

(ii) Why we have strong access to our mental states when we self-attribute them

In addition:

(iii) It must allow for the possibility that self-attributions of mental states are wrong.

Fernández argues, convincingly it seems to me, that alternative accounts of self-knowledge either satisfy (i) and (ii) at the cost of satisfying (iii), or satisfy (iii) at the cost of failing to satisfy (i) or (ii). In contrast, argues Fernández, Bypass satisfies all of (i), (ii) and (iii). Is that so?

Consider (ii). For (ii) to be true, Bypass must explain why our self-attributions of mental states are more secure than our attribution of mental states to others. Here is Fernández explanation of (ii):

> The truth of the Strong Access principle can be explained in terms of liability to error. In order for you to be justified in believing that I have a belief, you typically need to observe my behavior (including my verbal behavior) and infer from it that I have the belief in question as the best explanation of your observations. There are some aspects of this procedure that make you liable to error in ways in which I am not. Your perceptual experiences of my behavior may turn out to be wrong. Also, you may make a mistake while you are performing the relevant inferences….Consider by contrast, my self-attribution of a particular belief. It is easy to see that it is not vulnerable to those types of error [Fernández (2013), p. 57].

So far so good. When we attribute mental states to ourselves we do not seem to rely on making inferences or observation of behavior.

There is a familiar asymmetry between the basis for a self-attribution of a mental state and the basis for its attribution to another. What is hard to see is why, far from confirming that asymmetry, Bypass is not at variance with it. Suppose that the mental state I am self- attributing is my belief that you are in pain. According to Bypass my basis for self-attributing that mental state is the very same as my basis for attributing pain to you. So, my justification for believing that I believe you are in pain is your acting as though you are in pain. Hence, the self-attribution of the belief that you are in pain is epistemically mediated by the very same pain behavior as my attribution of pain to you. Why then think that less can go wrong when I self-attribute my second-order belief. True, I am using the same grounds to justify beliefs about different things: one your being in pain, the other my believing that you are. Still, the grounds are the same. What is left unexplained is why those grounds provide greater epistemic security when they are grounds for a self-attribution.

III

Let us consider how Fernández utilizes Bypass to solve three philosophical problems. The problems are posed by Moore's Paradox, thought insertion, and self-deception. Fernández, in my view rightly, takes the problem posed by Moore's paradox to be the problem of showing why it is irrational to hold a belief in an instance of:

NB: P, but I do not believe P,

or:

BN: P, but I believe not-P.

Invoking Bypass, his solution goes like this. Consider the following principle:

No Grounds
For any proposition P and subject S:
S should not believe that P if, all things considered, S finds no grounds for believing that P.

Suppose Samantha believes the following instance of NB:

> M: Moore was a philosopher, but I do not believe that Moore was a philosopher.

Samantha's reason for believing the second conjunct of M is that she has no reason to believe that Moore was a philosopher. But, she does believe that Moore was a philosopher. So, she holds a belief that she believes she has no reason to hold. Hence, according to *No Grounds,* she is irrational.

I may have missed something in this reconstruction of Fernández' diagnosis of Samantha's irrationality in believing M. But, if it is accurate, notice that Bypass plays no role in the diagnosis. At least, that is so if Bypass is understood as the view that the grounds for self-attributing a belief are the same as the grounds for holding it. What seems to be doing all the work in Fernández' diagnosis is *No Grounds.* But, *No Grounds* is quite independent of Bypass. *No Grounds* says nothing about the kind of reason one will have for attributing, or refraining from attributing, a belief. An introspectionist could accept *No Grounds*. Here is what such an introspectionist can say. If, when she exercises her quasi-perceptual faculty of intuition, Samantha detects no beliefs that qualify as grounds for believing that P, she should not believe that P.

In any case it is not clear that the diagnosis of Samantha's irrationality based on *No Grounds* succeeds. When Samantha believes that Moore was a philosopher, she may well have a reason to do so. Still, if she can fail to recognize that she believes that Moore was a philosopher, she can fail to recognize that she has the belief constituting her reason for believing that Moore was a philosopher. If so, she can fail to recognize that she has that reason when it functions as a reason for believing that she believes that Moore is a philosopher. Should she do so, she will not be violating *No Grounds.* The reason she has for her first-order belief does double duty as a reason for her second-order one. It is just that she does not find her grounds for believing that she believes Moore was a philosopher, and so, without any irrationality, believes the second conjunct of M.

In Chapter 5 Fernández applies Bypass to the puzzling phenomenon of thought insertion. An individual experiencing thought insertion is apt to maintain that some consciously held, first personally entertained, thought is not her own. Fernández asks what is different about

such an individual's experience of an 'inserted' thought which allows her to attribute it to another. His answer is, to my mind, a very plausible one. It is that the individual suffering from thought insertion fails to endorse the inserted thought. As Fernández puts it she fails to entertain the thought in question assertively despite its being consciously held.

What is it to entertain a belief assertively? With one qualification, it is, I should say, to be prepared to move from:

I believe that P,

to:

P.[1]

So, what is supposed to be the connection between entertaining a thought assertively and Bypass? In Fernández' view it seems to be this. Suppose, I attribute to myself the belief that Hilary Clinton will be the next President. According to Bypass my grounds for doing so are the very same as the grounds for believing that Hilary Clinton will be the next President. So, the self-attribution of the belief about Hilary Clinton makes salient to me the grounds I have for believing that she will be the next President, and, hence, disposes me to believe that she will be.

Here is one worry I have with this Bypass account of what it is to entertain a thought assertively. Why should the order, first- or second, of my belief make a difference to my awareness of the grounds for holding that belief? If I am not aware of those grounds when they are grounds for one of my first-order beliefs, why should their being grounds for a second-order belief increase my awareness of them? Something seems to be missing from the Bypass account.

Fernández is concerned to defend a transparency account of self-knowledge. One problem I find with his view of the assertiveness of thought stems from an ambiguity in the use of 'Transparency' as a label for the knowledgeable/justifiable self-attribution of belief. One use of 'Transparency' is introduced in the following much quoted passage from Gareth Evans:

> The crucial point is the one I have italicized: in 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 pre-

> cisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' I get myself 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), pp. 225].

Another, albeit closely related, use of 'Transparency' applies to the view stated in the following earlier passage by Roy Edgley:

> [My]own present thinking, in contrast to the thinking of others, is transparent in the sense that I cannot distinguish the question "Do I think that P?" from a question in which there is no essential reference to myself or my belief, namely "Is it the case that P?" This does not of course mean that the correct answers to these two questions must be the same; only that I cannot distinguish them, for in giving my answer to the question "Do I think that P?" I also give my answer, more or less tentative, to the question "Is it the case that P? [Edgley (1969), p. 90].

On the view stated by Edgely, from a first person perspective, I am entitled to move from: P to I believe P, and from: I believe P to P. Let us stipulatively reserve 'Transparency' as a label for that view. Transparency can be found in the passage from Evans. But, a view weaker than Bypass, call it Weak Bypass, can also be found towards the end of that passage. According to Weak Bypass if I have grounds for believing that P, then I have grounds [which may not be the same grounds] for believing that I believe P.

It is, I should say, Transparency rather than Bypass, or Weak Bypass that is relevant to entertaining a thought assertively. If I am prepared, within a first person perspective, to move from: I believe that P, to: P, then I am entertaining the thought that P assertively.

Fernández might reply that what sustains the move from: I believe that P, to: P, is finding grounds for believing that P. If he does, that takes us back to the Bypass account of what it is to recognize one's grounds for holding a belief. But, we have found that account wanting.

The last puzzle Fernández applies the Bypass account to is the one posed by self-deception. For Fernández self-deception is a failure of a certain kind of self-knowledge. So understanding it, he claims, helps us to explain two of its features, which he calls 'conflict' and 'normativity'. Conflict is that a self-deceived individual's pronounce-

ment about her belief conflicts with her other behavior. In one of the cases that Fernández describes, self-deceived Jack behaves in a way indicating he is sick while denying that he has that belief. Normativity is that we criticize the self-deceived person for being self-deceived.

Fernández accounts for the conflict aspect of self-deception by taking it to be the following failure of self-knowledge. The self-deceived individual fails to believe that she has a certain belief. Moreover, that failure results from failing to transfer the reasons she has for holding the first-order belief to function as reasons for holding the second-order belief that she has that first-order belief. Jack believes, and has reason to believe, that he is sick. He refrains from employing the reasons he has to believe he is sick to believe that he has that belief.

How does Fernández handle the normativity of self-deception? By taking Jack to violate a principle that we encountered already in the discussion of Moore's Paradox:

> *No Grounds*
> For any proposition P and subject S:
> S should not believe that P if, all things considered, S finds no grounds for believing that P.

So, how does Jack violate *No Grounds*? Jack believes that he is sick. But, he believes he does not have that belief. If he has grounds for believing that he lacks the first-order belief, according to Fernández, they will be the absence of grounds for the first-order belief that he is sick. So, Jack believes he is sick despite finding no grounds for doing so: thus violating *No Grounds.*

There are a number of things to be said about this account of the conflict and normativity involved in self-deception. I will confine myself to the following. As with the employment of *No Grounds* in connection with Moore's Paradox its employment in giving an account of the conflict and normativity associated with self-deception would seem to make Bypass redundant. To explain conflict we postulate a failure of self-knowledge. Jack erroneously believes he does not believe he is sick. But he does believe he is sick. So, what he says about holding that belief conflicts with how he behaves as a result of holding it. When we ask how the relevant failure of self-knowledge arises Bypass is supposed to supply the answer. But, again, it is open to the introspectionist to accept that Jack's self-deception arises from his forming the belief

that he does not believe that he is sick in an epistemically culpable fashion. Given *No Grounds* Jack is at fault in forming his second-order belief because he fails to properly deploy his introspective faculty of belief detection to detect both his first-order belief and the grounds for holding it. After all we are used to individuals misperceiving what is in front of them.

I have been quite critical of Fernández's book. I would not like to leave the reader with the impression that I do not think it is a very good book. It is an excellent contribution to the burgeoning literature on self-knowledge: one that cannot be ignored by anyone interested in the topic.

Philosophy Department
Syracuse University
Syracuse, NY 13244, USA
E-mail: agallois@syr.edu

NOTES

[1] The qualification I have in mind is this. Suppose someone asks me whether Hilary Clinton will be the next President. I answer: I believe so, but I am not sure. In giving this answer I am attempting to answer a question about Hilary Clinton rather than my state of mind. Despite that I am not prepared to move from: I believe Hilary Clinton will be the next President, to: Hilary Clinton will be the next President.

REFERENCES

EDGLEY, R. (1969), *Reason in Theory and Practice,* London, Hutchinson.
EVANS, G. (1982), *The Varieties of Reference*, New York, Oxford University Press.
FERNÁNDEZ, J. (2013), *Transparent Minds: A Study in Self-Knowledge*, Oxford, Oxford University Press.

RESUMEN

Se ofrece una crítica de la explicación del autoconocimiento de Jordi Fernández. Según esa explicación, se tiene conocimiento privilegiado de las propias actitudes proposicionales conscientes porque se usan las razones para tener esas actitudes como razones para su auto-adscripción. En este artículo se argumenta que la explicación de Fernández es plausible sólo si se invoca una concepción implausible de la justifica-

ción. Asimismo, se da un argumento para mostrar que, incluso si consigue justificar las auto-adscripciones en cuestión, la explicación de Fernández no puede proporcionar la justificación más básica de dichas auto-adscripciones. También se argumenta que no puede proporcionar ninguna justificación a favor de tales auto-adscripciones. Finalmente, se defiende que la explicación de Fernández no consigue dar una solución que no sea redundante a los tres problemas filosóficos a los que se aplica.

PALABRAS CLAVE: *transparencia, introspección, justificación, acceso privilegiado*.

ABSTRACT

I discuss Jordi Fernández's defence of his account of self-knowledge which he calls Bypass. According to Bypass I have privileged knowledge of my consciously held propositional attitudes by employing the reasons for having such an attitude as reasons for self-attributing it. I argue that Bypass is only plausible given an implausible story about justification that Fernández invokes to defend it. I also give an argument to show that, even if it succeeds in justifying the relevant self-attributions, Bypass cannot give the most fundamental justification of those self-attributions. I also argue that it cannot give any justification for them. Finally I argue that Bypass fails to give a non-redundant solution to three philosophical problems Fernández applies it to.

KEYWORDS: *Transparency, Introspection, Justification, Privileged Access*.