

**teorema**

Vol. XXXIX/1, 2020, pp. 25-40

ISSN: 0210-1602

[BIBLID 0210-1602 (2019) 39:1; pp. 25-40]

## Fernández on Transparency: Is the Bypass Procedure Compatible with Changes in Belief-Formation?

Martin Fricke

RESUMEN

Según Fernández, autoatribuimos creencias sobre la base de sus fundamentos, “pasando por alto” (*bypassing*) las creencias mismas que nos atribuimos. Mi artículo argumenta que este procedimiento se topa con problemas normativos y metafísicos cuando ocurren ciertos cambios en las maneras en que el sujeto forma sus creencias. Si el cambio es accidental, el problema es normativo: no se puede justificar la autoatribución de la creencia a través del procedimiento de *Bypass*. El problema metafísico consiste en que no está claro cómo el procedimiento puede reflejar cambio alguno en la formación de creencias, considerando que supuestamente sólo tome en cuenta los fundamentos (cambiantes) de las creencias, no las creencias mismas.

PALABRAS CLAVE: *transparencia, autoconocimiento, procedimiento de Bypass, formación de creencias, Jordi Fernández, Alex Byrne, Gareth Evans, Josep Prades.*

ABSTRACT

According to Fernández, we self-attribute beliefs on the basis of their grounds, “bypassing” the beliefs to be attributed. My paper argues that this procedure runs into normative and metaphysical problems if certain changes in the subject’s ways of forming beliefs occur. If the change is accidental, the problem is normative: self-attributing the resulting belief by way of *Bypass* cannot be justified. The metaphysical problem is that it is unclear how the procedure can reflect any change in belief-formation at all, given that it is not supposed to take into account anything but the (changing) grounds of the beliefs, not the beliefs themselves.

KEYWORDS: *Transparency, Self-Knowledge, Bypass Procedure, Belief-Formation, Jordi Fernández, Alex Byrne, Gareth Evans, Josep Prades.*

The purpose of this paper is to develop a specific objection to Jordi Fernández’s transparency account of the knowledge we have of our own beliefs. The objection is that the account cannot explain how we can

know our beliefs when our ways of forming them changes. This amplifies a point made by Josep Prades in a discussion of Fernández's book, *Transparent Minds*, that was published in *Teorema*, 34.1. In what follows, I will briefly sketch the problem Fernández aims to solve and what his proposed solution is. Then I shall present my objection, which has two parts: the first one relates to the normative side of Fernández's account, the second to its metaphysical side. My conclusion will be that his theory is closer to a related transparency account – that of Alex Byrne (2005, 2011) – than at first apparent.

#### I. THE PROBLEM OF SELF-KNOWLEDGE, ACCORDING TO FERNÁNDEZ, AND HOW *BYPASS* IS SUPPOSED TO SOLVE IT

According to Fernández, self-knowledge regarding one's own beliefs<sup>1</sup> has two important features that need to be explained, namely what he calls special access and strong access. We have a *special access* to our own beliefs because we do not have to rely on reasoning or on behavioural evidence to know them. This makes our access special in comparison to the access we have to the beliefs of other people. We can only know their beliefs by observing their behaviour (which includes their verbal behaviour) and drawing our conclusion, i.e. reasoning, from what we observe. We *can* also approach our own beliefs in this way, “from the outside”; but normally we do not. We know them (and no-one else's beliefs) in a more direct way.

According to Fernández, we also have *strong access* to our own beliefs. Strong access is constituted by the fact that our justification for believing that we have a certain belief is normally especially strong compared to the justification that others can have for believing that we have this belief. It might be said that we are less likely to go wrong when our own belief is in question than when dealing with someone else's belief. Again, exceptions might be possible, such as when we are self-deceived about a belief; but the normal case, according to Fernández, is that access to one's own beliefs is better justified than that to others' beliefs.

The combination of special and strong access is what Fernández calls *privileged access*. The “problem of self-knowledge,” as conceived of by him, “is the problem of explaining how we can have privileged access to our mental states” [Fernández (2013), pp. 7].<sup>2</sup>

The solution suggested by Fernández is a transparency account of self-knowledge regarding beliefs. Such accounts are inspired by Gareth

Evans's observation that "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]. On this view, it is by directing our attention outwards, at the world, that we can find out what we believe. We do not have to introspect inner items in our minds. Fernández endorses this idea and explains it thus: Our lower-order beliefs are based on other mental states, their grounds. The same grounds regularly give rise to the same beliefs in us. When I perceive an apple in front of me, I normally form the first-order belief that there is an apple in front of me. Here the perception of the apple is the mental state which serves as grounds for the first-order belief. Now, given that I *regularly* form the same first-order belief on the same grounds, Fernández notes that I can also form a corresponding higher-order belief on the basis of the same grounds, namely the belief that I have the first-order belief. Thus, when I perceive an apple in front of me, I cannot only form the first-order belief that there is an apple in front of me on the basis of this perception, but also the higher-order belief that I believe that there is an apple in front of me. The perception indicates what is the case in the world, i.e. what I perceive, and it normally also indicates that I believe what I perceive. As Fernández says, it has a "double role" of grounding the first-order belief and the second-order belief.

Fernández calls this procedure of forming higher-order beliefs on the basis of the same grounds as the lower-order beliefs self-attributed in them *Bypass*, because it is a procedure that "bypasses" the lower-order belief itself and instead bases the higher-order belief directly on the grounds of the lower-order belief. Fernández claims that we normally come to our higher-order beliefs through *Bypass*.

Suppose that this is correct. How does *Bypass* help to explain the special and strong access we enjoy in self-knowledge regarding beliefs?

Access to our own beliefs through *Bypass* is special because the procedure does not involve making any observations or inferences over and above those that are necessary for coming to have the lower-order belief attributed by way of *Bypass*. The subject does not have to observe herself or her beliefs nor does she have to draw inferences from such observations. Rather, the higher-order belief is formed directly on the same basis as the lower-order belief.

Fernández points out that the resulting self-attribution of belief is strongly justified in so far as it does not rely on veridical perception or correct inferences. Of course, the truth of the *lower-order* belief might rely

on perceiving correctly or on drawing valid inferences. But the truth of the higher-order belief does not. Its truth depends solely on the subject correctly using the basis of the lower-order belief in order to attribute this same belief in a higher-order one. The self-attribution not depending on truthful perception or inferences, and therefore not being subject to failure in these capacities, is the reason Fernández regards it as providing strong access to one's own beliefs.

There is much more to be said about Fernández's theory and he develops it with exceptional clarity, care and thoroughness. However, for the purposes of this article, this rough sketch of his theory about self-knowledge regarding belief suffices. In what follows, my main question will be whether the *Bypass* procedure can cope with changes that might occur in our ways of forming beliefs. In such a case, will the procedure still produce strongly justified self-attributions of belief (the normative question)? And will it still be functional in the sense of producing correct self-attributions at all (the metaphysical question)?

## II. THE NORMATIVE FAILURE OF *BYPASS* WHEN OUR LOWER-ORDER BELIEFS ARE ACCIDENTAL

Josep Prades, in his discussion of Fernández's book, invites us to consider someone who sees a stick in water and comes to believe that the stick is straight, although it visually appears to be bent. According to Fernández, the subject can know through *Bypass* that he believes that the stick is straight, because he regularly forms the lower-order belief (that the stick is straight) given the kind of perception he has and given his background belief that watching a stick in water produces certain visual illusions. His perception and his background beliefs constitute the grounds on the basis of which he regularly forms his lower-order belief and thus can also serve as basis for ascribing the higher-order belief. Prades attempts to cast doubt on this application of *Bypass*; but I shall not discuss this point here. I *am* interested in his further complication of the example: "suppose that our subject is very nervous, or has taken some drugs, or has not slept in the last 48 hours. As a result of his anomalous psychological condition, he has the psychological disposition to trust the visual appearance of the stick, and he forms the belief that the stick he is seeing is bent. This would be an abnormal belief-formation process..." [Prades (2015), pp. 113].

Prades does not say so, but implies that the *Bypass* procedure would fail in such abnormal situations in the sense that it would not be able to provide a justification for the attribution of the lower-order beliefs. The resulting self-attributions of belief could not count as knowledge. This – Prades suggests – is implausible, as such cases do “not seem [to] involve a deficiency in self-knowledge at all” [cf. *ibid.*, p. 114].

Prades’s case is interesting because it deals with what Fernández, in his response to Prades, calls a “fluke error”; that is, an error that occurs accidentally, not habitually. A subject might be prone to habitual errors. She might not know about the optical illusions caused by seeing through water. In consequence, she will *regularly* judge the stick in water to be bent, although in reality it is not. Similarly, a subject might *regularly* reason fallaciously. She might tend to conclude from “If p, then q” and “q” that p, for example. These are cases where the subject’s lower-order beliefs are regularly formed in a faulty way on the basis of grounds that do not actually justify them. (That Fernández still uses the notion of “grounds” in these cases shows that for him the term does not carry any normative connotations and is only characterised by its causal role. [Cf. Fernández (2013), p. 45, footnote 7.]) Beliefs based on grounds that do not justify are likely to be false. However, as Fernández points out, this does not mean that they cannot be self-attributed using the *Bypass* procedure. Since the subject *regularly* reasons in the fallacious way, *Bypass* is justified and the subject is justified in forming the belief that she believes that the stick is bent on the basis of her visual perception; or the belief that she believes that p on the basis of the apparent facts that “If p, then q” and “q”. Habitual error in the formation of lower-order beliefs does not preclude strong access to them via *Bypass*.

However, a fluke error in the formation of a lower-order belief is not a habitual error. Rather, it occurs accidentally in virtue of the special conditions the subject finds herself in. It is just a temporary change in the subject’s ways of forming beliefs. This means that there is no regularity in the relation between the initial mental state and the lower-order belief formed on the basis of it.<sup>3</sup> The mental state does not *tend* to cause the belief in the subject; it does so only on this special occasion. In consequence, following the *Bypass* procedure and self-attributing the belief on the basis of the mental state that irregularly caused it on this occasion is *not* justified. The justification of *Bypass* depends on the regularity of the relation between the initial mental state and the lower-order belief. Since there is no such regularity when the belief is the result of a fluke error,

the self-attribution resulting from *Bypass* lacks a justification. This is the reason why, by Fernández's own lights, it should not count as self-*knowledge*.<sup>4</sup>

Note that the above argument can also be made by appeal to the occurrence of beliefs that are accidentally *true*. Just as in the case of fluke error beliefs, accidentally true beliefs are also formed irregularly, not habitually, on the basis of the states that are their grounds. In consequence, self-attributing them via *Bypass* would lack justification and the resulting higher-order beliefs should not count as self-*knowledge*. In what follows, I shall concentrate on the case of fluke error beliefs; but the argument can be run analogously for fluke true beliefs.

Is it plausible to deny that we can have the same kind of privileged self-knowledge about our accidentally formed beliefs as we do about those that are not accidental? Why should this not be possible? If we think that self-knowledge about accidental lower-order beliefs should be possible in just the same way as self-knowledge regarding regularly formed beliefs, then we have to reject the *Bypass* theory.

Fernández responds to this objection that we should accept “that if I form some first-order belief by a total fluke, then I do not qualify as knowing that I have that belief” [Fernández (2015), pp. 150]. This is because he thinks that the conditions must be rather “seriously abnormal for me to lack any sort of grounds for my first-order belief”. But “if I have taken drugs, or I have not slept in the last 48 hours, and I am not responsible for the first-order beliefs that I am forming, then what reason is there to think that I will be competent in attributing beliefs to myself?” [Ibid.] Fernández seems to regard the conditions that bring about fluke errors as so out of the ordinary, and perhaps irrational, that an incapacity of self-knowledge might as well accompany them.

Although Fernández's response makes sense within the framework of his theory, it seems to me that it has a peculiar implication which might lead us to regard it as doubtful. The implication in question is that fluke errors and other accidental beliefs are rare phenomena that result from serious failures of rationality in the formation of beliefs. This implication is peculiar because clearly accidental errors occur all the time in our daily lives. I remember that my keys are in the drawer, but in fact they are not. I see someone at the door, but in fact there is no-one. I calculate what I will have to pay the waiter, but get the sum wrong. All these little accidental errors must be, for Fernández, errors that occur in the formation of states that ground our beliefs; not errors in the formation of the beliefs themselves on the basis of whatever their grounds

are. They must be errors of perception, not errors in the formation of belief on the basis of perception; errors of memory, not errors in the formation of belief on the basis of memory; errors in the process of forming the state on the basis of which I come to believe what the sum is, not errors in the forming of this belief itself. For otherwise, all these accidental errors would be cases in which self-knowledge through *Bypass* would seem to be impossible, since there is no regularity in the coming about of the error. If we doubt that daily accidental error can be located so clearly in the process of forming mental states that ground our beliefs, rather than in the formation of those beliefs themselves, then the kind of fluke error that puts into question the justification of our self-attributions of belief in Fernández's theory becomes much more common than he seems to allow.<sup>5</sup>

### III. THE METAPHYSICAL FAILURE OF *BYPASS*: IT CANNOT IMMEDIATELY REFLECT CHANGES IN THE SUBJECT'S WAYS OF FORMING BELIEFS

In the previous section we have considered the normative aspect of Fernández's transparency theory: Do the deliverances of *Bypass* constitute knowledge when the lower-order beliefs are accidentally erroneous? In this section I wish to consider a metaphysical aspect of the theory: Is *Bypass* capable of producing self-attributions at all (whether constituting knowledge or not) if the subject's ways of forming her lower-order beliefs change?

To discuss this question it is useful to reconsider the example of fallacious reasoning mentioned earlier: a subject regularly reasons according to the fallacy of affirming the consequent. Given premises such as "q" and "If p, then q", she regularly concludes that p. In this case, *Bypass* should provide her with knowledge that she believes that p, because the prior beliefs that q and that if p, then q are the grounds for her belief that p, i.e. they regularly cause her to believe that p. The regularity is not rational, but it suffices to justify the application of *Bypass*. Now let us suppose that at some point the subject stops reasoning according to this particular fallacy. It might be that she realises that her earlier reasoning was faulty. Or it might be for some other reason that she changes her conclusions. From now on, when she believes premises such as "q" and "If p, then q", she no longer concludes that p.

Now my question is this: how does the subject know, on the basis of her beliefs that  $q$  and that if  $p$ , then  $q$ , whether or not she believes that  $p$ ? According to *Bypass*, the higher-order belief “I believe that  $p$ ” is formed exclusively on the basis of the states that ground the subject’s belief that  $p$ . But while still reasoning fallaciously, the beliefs “ $q$ ” and “If  $p$ , then  $q$ ” ground (in the causal sense) the subject’s belief that  $p$ ; while later on they no longer do so. How is the subject to know which of the two possibilities is the case, on the basis of the prior beliefs “ $q$ ” and “If  $p$ , then  $q$ ” *alone*? Nothing in the beliefs themselves indicates whether or not the subject uses them as grounds for forming the belief that  $p$ . If we take seriously the idea that higher-order beliefs are formed *bypassing* the lower-order beliefs to be ascribed (and their formation processes), then there does not seem to be any way that the mental states prior to the lower-order beliefs can tell the subject whether or not they ground such lower-order beliefs. If we admit that the belief-formation might be irrational (though still regular), then it is possible that beliefs are formed on the basis of states that normally should have no propensity to bring them about. And when beliefs *are* formed rationally, it was still the case, prior to their formation, that the subject might as well not have formed them, but some others in an irrational way. So, it seems that the states on the basis of which beliefs are formed do not indicate, by themselves, whether those beliefs are formed by the subject on a given occasion, or which ones of them, those rationally implied or others, are formed. If this is correct, then Fernández’s theory fails to explain how it is possible by means of *Bypass* to arrive at correct self-attributions of belief.

What could Fernández reply to this argument? One response could be to say that the subject learns from experience whether or not the *Bypass* method can be used to self-attribute a belief on the basis of certain prior states. If the states – here: the beliefs “ $q$ ” and “If  $p$ , then  $q$ ” – regularly lead to the belief that  $p$ , then the subject might realise that self-attributing the belief that  $p$ , given that  $q$  and that if  $p$ , then  $q$ , normally leads to a truthful self-ascription of belief. And if the subject stops forming beliefs according to this fallacious scheme, she might again, after a while, learn from experience that the self-attribution of the belief will now turn out false, when made on the basis of the same states as before.

However, this view of how specific *Bypass* routines are formed does not seem very plausible. It would mean that at least in some cases the transparency of beliefs (i.e. that they can be accessed via *Bypass*) is the result of a prior third-personal access to them, because “learning from experience” what beliefs one forms in certain types of situation seems to

rely on some form of observing oneself as one would observe a third person. This view also seems to imply that we fairly often go wrong in our self-attributions of belief when we are still in the process of learning from experience what specific ways of using *Bypass* are correct. But such episodes of frequent uncertain or false self-attributions do not seem to exist in our lives. Another problem with this view is that it is unclear how general the *Bypass* routines that we would have to acquire can be. Perhaps we are more prone to affirm the consequent when specific subjects are under consideration. If this were so, a distinct learning process would seem to be necessary for self-attributing these beliefs as opposed to any others. As a result, whenever new subjects and perhaps new forms of reasoning are under consideration, the subject would have difficulty knowing about at least some of her own beliefs because she would lack the experience necessary for forming correct *Bypass* routines. Again, this does not seem very plausible. It does not seem to be the case that we lack knowledge of what we believe just because we believe it on the basis of inferences that are new to us.<sup>6</sup>

The metaphysical problem with the *Bypass* procedure is not restricted to self-knowledge with respect to fallaciously formed beliefs. It can also be developed with respect to beliefs formed on the basis of perceptual states. Suppose a subject sees greenery and forms the belief that there is a birch tree. Let us suppose also that the subject uses an established *Bypass* routine to form, on the grounds of her perceptual state (a visual experience that there is such-and-such greenery) and her classificatory belief (“Such-and-such greenery is a birch tree”), the higher-order belief that she *believes* that there is a birch tree. Now, there seem to be at least two ways in which her formation of lower-order beliefs might change without her *Bypass* procedure being able to take this into account. First, it might be that she does not make use of her classificatory beliefs because she is distracted. We might say that she does not think about what kind of greenery she perceives; her classificatory beliefs simply do not get involved with her perceptual state. Second, it might also be that some other mental state interferes with her formation of beliefs. For some reason she might come to believe that she does not remember correctly what birch trees look like, for example. This might have the effect that, although the classificatory belief gets activated in some way by her perceptual state, she still does not form the belief that there is a birch tree. The first case is somewhat similar to the fluke errors discussed earlier; the second case simply represents a situation in which the subject receives

new information and therefore no longer forms the same beliefs in response to the same perceptual state as before.

Would Fernández's *Bypass* procedure be responsive to these changes in the subject's ways of forming beliefs? The subject's perceptual state and her classificatory beliefs would be the same in the situation where she does form the belief that there is a birch tree as in the situation where she does not. There is nothing in these states – the compound state, as we might call it – to indicate whether or not the subject will form the belief that there is a birch tree. So how could the subject find out, using *Bypass*, whether or not she has this belief? Since *Bypass* is supposed to base self-attributions of belief exclusively on the mental states grounding the belief to be attributed, not on the process of belief-formation that might be initiated by these states, there is no clear way in which this procedure could reflect possible changes in the subject's ways of forming beliefs. If she has learned to self-attribute the belief that there is a birch tree on the grounds of a certain perceptual state and her classificatory beliefs, then it seems she should continue to do so, even if on a given occasion the subject fails to form the belief because she is distracted or because she no longer trusts her classificatory knowledge.

Fernández would probably reply that in the first case (distraction) the subject is making a fluke error and that the corresponding false self-attribution of belief might indeed occur; but this is not surprising given that the kind of distraction stipulated constitutes a significant impairment of the subject's rationality. The problem with this reply is that it is not very plausible. False self-attributions of belief seem to be a very rare phenomenon, probably rarer than simple failures of rationality such as not believing anything about the classificatory details of a perceived scene despite having general classificatory knowledge at one's disposal.

In the second case (an interfering state of distrust), Fernández might respond that the subject's perceptual state, her classificatory beliefs *and* her distrust in them, taken together, constitute a new compound state that normally is not apt to ground the formation of the belief that there is a birch tree. For this reason, the subject does not form the higher order belief that she believes that there is a birch tree. The problem with this response is, as I have tried to argue, that it is unclear how the subject can come to detect through *Bypass* whether a given compound state, say a perceptual state and a classificatory belief, functions as grounds (if there is no interfering state) or not (if there is such an interfering state). Given that *Bypass* does not seem to take into account the *process* which leads from the grounds to the formation of the lower order

belief, it is hard to see how it could be responsive to such changes in the subject's ways of forming beliefs on the basis of given states.

I shall conclude by suggesting how Fernández should reply to my objection. I believe he should say that the *Bypass* procedure works because it does not only take into account the *presence* of states that can ground subsequent beliefs, but also the “pull” these states exert on the subject to form beliefs. In other words, the procedure must also take into account at least part of the process by which the grounding states lead to the formation of beliefs. We might say that it must be responsive not only to the grounds but also to their “causal effects *downstream*” [Ashwell (2013)]. If a subject is inclined to reason fallaciously, then, on the basis of the fact that  $q$  and that if  $p$ , then  $q$ , it *seems* to her that  $p$ . If she is perceiving a birch tree, but distrusts her classificatory knowledge, then it does *not seem* to her, all things considered, that there is a birch tree.<sup>7</sup> *Bypass* must detect such overall seemings, the “pull” that our grounding states exert on our belief-formation. Only such seemings can indicate in which way our processes of belief-formation will go. Otherwise it must remain mysterious how a procedure such as *Bypass* should be able effortlessly to take into account changes in our ways of forming beliefs.

Is it problematic to modify Fernández's theory in the manner suggested? I think that it is not; rather, it makes the theory more plausible.<sup>8</sup> However, an interesting point to note is, I think, that when we compare Fernández's theory with other transparency theories of self-knowledge, the modification makes his theory rather more similar to Alex Byrne's than is at first apparent. Byrne suggests that we come to know our own beliefs by inferring them from what we (seem to) know about the world according to the following “doxastic schema”:

$p$

---

I believe that  $p$ .

[Byrne (2011), p. 204; Byrne adopts the schema from Gallois (1996), p. 46.]

Without going into the details of why Byrne thinks that following this schema can provide one with self-knowledge, it is immediately salient that, on his view, we form higher-order beliefs on the basis of the beliefs that those higher-order beliefs are about. We form the belief “I believe that  $p$ ” on the basis of the belief that  $p$ . On the face of it, this seems to

make Byrne's account rather different from that of Fernández, because, according to the latter, we should *bypass* the lower-order belief and form the higher-order belief on the basis of the *grounds* we have for the lower-order belief, not on the basis of this lower-order belief itself. Both accounts are transparency theories in that they claim that we find out what we believe by directing our attention outwards at the world, rather than inwards at our inner mental states. But on Fernández's view, this outward attention produces mental states that serve as grounds for our lower-order beliefs and, simultaneously, for our self-attributions of these beliefs. Byrne, on the other hand, claims that the outward attention produces first the lower-order beliefs and then, on the basis of them and by way of an inference according to the doxastic schema, our higher-order beliefs.

This sharp distinction between Fernández's account and that of Byrne becomes much less pronounced with the proposed modification of Fernández's theory. Now, it is no longer simply the mental state which grounds the subject's lower-order belief, but also the fact that this state "pulls" the subject towards the formation of this belief, which serves as basis for forming the higher-order belief. It seems to me that this modification reduces considerably the difference to Byrne's account. We might say that, according to Byrne, the subject (apparently<sup>9</sup>) *realises or notes that p* and consequently forms the belief that she believes that p; while according to Fernández, suitably modified, it *seems to the subject that p, all things considered*, and consequently she forms the belief that she believes that p. The *Bypass* procedure, in the modified account, seems to be very similar indeed to that described by Byrne's doxastic schema.<sup>10</sup>

#### IV. CONCLUSION

Fernández's transparency theory develops the ingenious idea that we come to know what we believe on the basis of the same mental states that ground those lower-order beliefs. Thus, our formation of higher-order beliefs "bypasses" the lower-order beliefs to be attributed. There are two problems with Fernández's conception of the *Bypass* procedure, which both have to do with the fact that the ways in which we come to our beliefs can change in time. First, there is a normative problem: self-attributions of one's own beliefs cannot be justified by *Bypass* if the beliefs to be attributed have been formed accidentally, i.e. not in our habitual way. Second, there is a metaphysical problem: it is unclear how *Bypass* could possibly reflect changes in one's way of forming beliefs, be they

accidental or real changes of mind. There is a remedy for the second problem: Fernández should say that we self-attribute beliefs not only on the basis of the mental states which are the grounds for the beliefs to be attributed, but also by taking into account the direction in which those grounds move us to form beliefs, the “pull” they exert on our belief-formation. I have argued that this modification of Fernández’s theory makes it very similar to Alex Byrne’s transparency account of self-knowledge.

*Instituto de Investigaciones Filosóficas  
Universidad Nacional Autónoma de México  
Circuito Maestro Mario de la Cueva s/n, Ciudad Universitaria  
C.P. 04510, Coyoacán, Ciudad de México, Mexico  
E-mail: mfcephcis@gmail.com*

*Departamento de Humanidades y Procesos Sociales  
Escuela Nacional de Estudios Superiores, Unidad Mérida  
Universidad Nacional Autónoma de México  
Carretera Mérida-Tetiz Km 4.5  
Municipio de Ucú, C.P. 97357, Yucatán, Mexico*

#### ACKNOWLEDGMENTS

This text was written while on sabbatical leave at Bielefeld University, Germany. I have presented earlier versions of some of the ideas at Instituto de Investigaciones Filosóficas, Universidad Nacional Autónoma de México; Universidad de Concepción, Chile; Bielefeld University, Germany; and at conferences in San José, Costa Rica; San Cristóbal, Mexico; Munich; Villa de Leyva, Colombia; and Umeå, Sweden. I would like to thank the audiences on all these occasions for their comments and criticisms. I am especially grateful for the thorough discussion my paper received at the Research Seminar organized by Christian Nimtz at Bielefeld University. I would also like to thank Javier Vidal, David Mathers and an anonymous referee of *Teorema* for helpful comments.

#### NOTES

<sup>1</sup> Fernández makes his observation with regard to knowledge of our own propositional attitudes in general [cf. Fernández (2013), pp. 4ff.]. Here I shall concentrate on the case of belief only.

<sup>2</sup> Fernández does not mention this point, but it seems to me that the urgency of explaining privileged access can be brought out by pointing out that, normally, stronger justification of empirical knowledge goes together with repeated observation, well-proven inference routines, back-up by theory or con-

sultation with other experts on the matter. But all these elements of justification are excluded by the “specialty” of access as defined by Fernández. From this perspective, having *only* special access or *only* strong access to some realm of facts looks less problematic than having both at once. The *simultaneity* of both in the privileged access we have to our own beliefs, then, is in particular need of explanation.

<sup>3</sup> As a referee points out, saying that a state *regularly* brings about a belief might be too demanding a claim compared to saying that it *reliably* brings it about. We might say that, in the first case, we are claiming that there are several occasions on which the state invariably brings about a belief of this kind, while in the second case, the state might cause a belief of this type only once, but reliably so, i.e. perhaps in such a way that this would normally happen in these circumstances. Fernández generally reserves the term “reliable” to characterise our capacities of perception, memory and reasoning or the testimony we depend on, while using “regularly” when speaking of the formation of beliefs [cf. e.g. Fernández (2013), p. 65]. But it seems clear that he cannot rule out the possibility of self-ascribing beliefs formed only once. So, it seems that “regularly” must mean more for him than just “on repeated occasions”. I think this is also fairly clear from the fact that he describes the regularity in first-order belief-formation by saying that a state *tends* to cause a certain belief [ibid., p. 45]. He might shun the term “reliable” in the context of belief-formation because he rejects a functionalist reliabilism that says that certain mental states under specific circumstances reliably produce beliefs to the effect that we have them. He thinks that such reliabilism fails to account for a sense of “cognitive achievement” associated with self-knowledge [cf. ibid., p. 26ff.]. The *Bypass* procedure is meant to preserve this sense by requiring that the subject concentrate on what grounds she has for her lower-order beliefs in order to self-attribute them [cf. ibid., pp. 62f.].

<sup>4</sup> Strictly speaking, the *Bypass* procedure does not have application in the case of fluke error beliefs. This is because of the following two facts: (1) *Bypass* as defined by Fernández requires that the higher-order belief is formed on the basis of a mental state that constitutes *grounds* for the lower-order belief to be self-attributed [cf. Fernández (2013), p. 49]. (2) Fernández defines “mental states constituting grounds for some belief in a subject” as states which make it the case that the subject *tends* to have the belief when being in the state [cf. Fernández (2013), p. 45]. Clearly, in the case of fluke error the mental state on the basis of which the erroneous belief is formed does not *tend* to cause the belief. It only does so on this abnormal occasion. So strictly speaking, fluke error beliefs *do not have grounds*, according to Fernández, and thus there is no mental state constituting grounds for the fluke error belief on which the *Bypass* procedure could base the second-order belief.

<sup>5</sup> One way of questioning the rarity of accidental error in belief-formation is to question Fernández’s idea that we always form our lower-order beliefs on the basis of other *mental states*. Is it necessary to stipulate that there always has to be an intermediate state in between the world and our beliefs? Why cannot be-

liefs be formed *directly* on the basis of those states of affairs in the world that they are about, without an intermediary state? (Belief theories of perception come to mind [cf. Armstrong (1968)].) If this were a common way of forming lower-order beliefs and if accidental error frequently occurred in such beliefs, as our daily experience suggests it does, then it would again be impossible to keep the error separate from the belief-formation and self-attributions of belief through *Bypass* (suitably reformulated) would lack justification in such cases.

<sup>6</sup> A second reply Fernández might give to the earlier argument is that he construes inferential reasoning in such a way that the problem cannot arise. His description of the regularity involved in inferential reasoning is this: “If S believes that Q and S believes that P follows from Q, then S comes to believe that P” [Fernández (2013), p. 46]. If we apply this scheme to someone reasoning according to the fallacy of affirming the consequent we get that this subject believes that q and that if p, then q and *additionally* also believes that it follows from these facts (*viz.* “q” and “if p, then q”) that p. Fernández might say that in this description there is an element in the subject’s prior states which shows that she will form the belief that p, namely her additional belief that p follows from the other two beliefs. – I think there are three problems with this view: (i) Fallacious reasoning, on this view, does not seem to be really fallacious. With the additional belief, taken as a premise, the conclusion *does* follow from the premises; the fallacy is reduced to a false premise. (ii) There is a threat of an infinite regress: if in order to come to believe that P on the basis of the subject’s belief that Q she has to believe that P follows from Q, it seems that we are required to add infinitely further beliefs to the effect that P follows from the previous premises. (iii) Even if we accept Fernández’s reconstruction of fallacious reasoning, it is still unclear how the subject can know, on the basis of her prior beliefs, which are now three (“q”, “If p, then q”, “p follows from q and from the fact that if p, then q”) that she believes that p. Since Fernández concedes that belief-formation can be irrational and change over time, there is no guarantee that these three beliefs will make the subject conclude that p, rather than something else or nothing at all.

<sup>7</sup> Fernández seems to suggest at one point that having grounds for the belief that p can be identified with it seeming to the subject that p [cf. Fernández (2013), pp. 47f.]. I have two remarks: first, although Fernández argues otherwise (*ibid.*), there is very little difference between it seeming to someone that p, all things considered, and her believing that p. Second, a given state, say a perception, might result in an overall seeming or it might not, if some other state interferes. So, it seems reasonable to regard the state and the seeming as distinct states.

<sup>8</sup> Does the modification also help Fernández reply to my first (normative) objection? He might say that even fluke errors are reliably based on some kind of overall “seeming” and that therefore the corresponding self-attributions of belief can also be based on it. This raises the question of why we need the state prior to the seemings (in unflukey situations) at all.

<sup>9</sup> Realising (like noting) is a factive verb; so realising that *p* presupposes that it is true that *p*. Here it might be that the subject only appears to realise that *p*, because *p* is not the case. Byrne's doxastic schema would still produce a correct self-ascription of *belief*. Byrne describes this feature of the schema by saying that it is "strongly self-verifying" [Byrne (2011), p. 206].

<sup>10</sup> One remaining difference between Fernández and Byrne might be that only the former clearly assumes that all beliefs are formed on the basis of prior *mental* states such as perceptions. Although Byrne does not defend a view of perception such as Armstrong's, which reduces perception to belief, he does think that "vision constitutively involves belief?" [Byrne (2012), p. 205]. It might be that this idea is incompatible with Fernández's claim that such belief is *caused* by perceptual states; but the point would need to be argued in more detail.

#### REFERENCES

- ARMSTRONG, D. M. (1968), *A Materialist Theory of the Mind*, London, Routledge and Kegan Paul.
- ASHWELL, L. (2013), 'Review of *Transparent Minds: A Study of Self-Knowledge* by Jordi Fernández', *Notre Dame Philosophical Review*, vol. 8, <https://ndpr.nd.edu/news/transparent-minds-a-study-of-self-knowledge/> (consulted 8/02/2014).
- BYRNE, A. (2005), 'Introspection', *Philosophical Topics*, vol. 33, pp. 79-104.
- (2011), 'Transparency, Belief, Intention', *Proceedings of the Aristotelian Society Supplementary*, vol. 85, pp. 201-21.
- (2012), 'Knowing What I See', in Smithies, D. and Stoljar, D. (eds.), *Introspection and Consciousness*, Oxford, Oxford University Press, pp. 183-209.
- EVANS, G. (1982), *The Varieties of Reference*, Oxford, Clarendon.
- FERNÁNDEZ, J. (2013), *Transparent Minds. A Study of Self-Knowledge*, Oxford, Oxford University Press.
- (2015), 'Replies to my Critics', *Teorema*, vol. 34, pp. 149-60.
- GALLOIS, A. (1996), *The World Without, the Mind Within. An Essay on First-Person Authority*, Cambridge, Cambridge University Press.
- PRADES, J.L. (2015), 'Transparency, Justification and the Achievements of Self-Knowledge', *Teorema*, vol. 34, pp. 107-19.