JORDI FERNÁNDEZ. 2013. *Transparent Minds. A Study of Self-Knowledge.* Oxford: Oxford University Press.

Gareth Evans famously claimed that "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 precisely the same outward phenomena as I would attend to if I were answering the question 'Will there be a third world war?' (1982, 225). Jordi Fernández' *Transparent Minds* aims to develop this suggestion. Here I will not be able to do it full justice, given the space allotted and will confine myself to a presentation and discussion only of some of its central claims.

In the first chapter, Fernández presents the problem of self-knowledge and discusses some rival accounts. Any explanation of self-knowledge, claims Fernández, should explicate how come that we have "strong access" to our mental states. That is to say, it should explain how we can have knowledge of them other than through the observation of our outward behavior and inference to the best explanation; and it should explain why our psychological self-ascriptions are more likely to be correct than the corresponding third-personal ones.

This apparently innocuous formulation in fact proposes a weak reading of the idea that we have privileged access to our own mental states, by building into the notion the contention that the difference between a first- and a third-person perspective onto our own mental states depends merely on relying, or not, onto *behavioral* criteria and on inferences to the best explanation, when in fact also the introspectionist account of self-knowledge should be considered. Furthermore, it builds into the notion of strong access the idea that the especially secure way in which we know our own mental states differs from other people's knowledge of them only as a matter of *degree* rather than in kind, for it is merely more secure than its third-personal counterpart. However, neither assumption is water-tight and many theorists would dispute both of them.

Fernández then identifies five further desiderata for a theory of self-knowledge. They are: the requirement, which many theorists in fact don't share, that self-knowledge be based on some form of cognitive achievement; the idea that cases of wrong self-ascriptions must be allowed; the requisite that the transparency of our mental states must be respected; the constraint that an explanation is owed as to why Moore-paradoxical contents are irrational; and, finally, the requirement that an account be given as to of why our self-ascriptions of propositional attitudes are "assertive", i.e. why they "put pressure" on us to occupy the corresponding first-order mental states.

In the second chapter, Fernández presents the bypass model for belief. The central thesis is that one's psychological self-ascriptions are based on whatever grounds support the corresponding first-order beliefs (p. 49). So the former are at least prima facie justified if one's beliefs are as well, and usually they are, if they follow the "production-of-belief principle" (p. 46 and p. 55): if one apparently perceives/ remembers/ acquires testimonial information/ has an intellectual seeming that P, one comes to believe P; and so one does if one believes Q, and believes that Q entails P.

Fernández then addresses the objection that the bypass model is a better explanation of how one creates one's beliefs rather than of our knowledge of pre-existing ones. He concedes that many times that is the case, but he contends that there is also a non-deflationist reading of the question "Do you believe that P?" whose answer depends on the application of the bypass model. I must confess that I wasn't persuaded by the argument offered in defense of the bypass model, but I won't dwell on that.

Fernández then explains how the bypass model satisfies the privileged-access and the strong-access requirement. It does so because we enjoy certain mental states (like experiences, memories, etc.) which naturally and usually lead us to form the corresponding beliefs as well as second-order self-ascriptions, without any need to observe our own overt behavior and without having to reason from the basing mental state to the second-order self-ascription (p. 57). By contrast, other people will be in a position to ascribe a belief to us only by observing our overt behavior and by inferring to its likely cause. If, however, our self-ascriptions aren't based on outer observation and inference to the best explanation, they aren't open to the kinds of error that can beset third-personal ascriptions of belief. Hence, our own access to our own beliefs is different and also more secure than whatever access other people might have with respect to them. Fernández also claims that the bypass model is not causal, because the self-ascription is not causally brought about by the first-order belief, but, rather, it is grounded on the evidence on which that very belief is based.

However, it seems odd to hold that one's self-ascriptions of belief should be justified by the evidence that prompts the first-order belief itself. Intuitively, the self-ascription, if justified at all, should be justified by the corresponding first-order mental state. This would indeed match the intuition that while we may have unjustified first-order beliefs, if they were formed on inappropriate, scant or no grounds, we could nevertheless be justified in self-ascribing them, for the simple reason that we do have them. To rescue this compelling intuition, Fernández has to say (p. 66) that while the subject would have evidence but no justification for the first-order belief, that very evidence would count as a justification for her self-ascription. Now, besides the difficulty of distinguishing between evidence and grounds in a satisfactory way, it is clear that this could be so only if we considered the role of that evidence in giving rise to the first-order belief which would then tend to correlate with one's self-ascription of it. But, if that is the case, in the end what "justifies" one in self-ascribing the belief is precisely the first-order belief itself, not the evidence that prompts it.

Moreover, it is difficult to see how the bypass model would count as non-causal. For, after all, it depends on the fact that once we are exposed to certain kinds of evidence, we tend to form the corresponding first-order beliefs and, if conceptually equipped, we also tend to form the relevant psychological self-ascriptions (p. 61). True, if Fernández' story were right, the causal basis of one's psychological self-ascription wouldn't be the first-order belief but the grounds on which the latter is based, but the eventual psychological self-ascription would be arrived at by means of a causal process nonetheless.

Afterwards, Fernández considers two further objections to the bypass model: the one from the absence of grounds and the one from the absence of belief. The former

has it that if one has lost the evidence on which a given first-order belief of hers is based, this would seem to entail, on the bypass model, that she couldn't self-ascribe that belief. The latter, in contrast, points out that the bypass model has difficulties in accounting for our own knowledge of lack of belief. The response to the former objection is that as long as one seems to remember that P, no matter what evidence originally prompted one's belief that P, one has evidence for the self-ascription of the belief that P. The answer to the latter, in contrast, is that by reflecting on P one can find out that there are no sufficient grounds to endorse it (or its negation) and therefore form the belief that one does not believe that P.

These replies seem to me to miss the point of the objections. For the first hinges on the idea that it cannot be the evidence *that led one to believe that P* that justifies one's self-ascription. The latter, in contrast, hinges on the fact that while there are no sufficient grounds to believe that P, there is sufficient ground to self-ascribe the absence of belief. So the grounds on which these beliefs are based cannot be one and the same.

Fernández claims that his account satisfies the desideratum that self-knowledge be a kind of cognitive achievement for it takes some work to gather and assess the evidence in favor of P. Yet here Fernández seems to be re-defining the requirement itself. For when theorists debate over the issue of whether self-knowledge depends on some kind of cognitive achievement, they all take themselves to be disputing over whether there is a viable sense in which, *by having in view a first-order belief*, we come to form the corresponding self-ascription. Now, on that understanding of the desideratum, Fernández' proposal would actually turn out to deny that self-knowledge is the result of some cognitive achievement.

As to transparency, it should be noted that there is a traditional understanding, according to which one's first-order propositional attitudes are immediately known to one (provided one has the relevant conceptual repertoire). Their occurrence, therefore, is of a piece with one's knowledge of them. By contrast, following a more recent use of that term, Fernández means the idea that in making the psychological self-ascription we literally bypass the first-order mental state to look only at the world, so to speak. Fernández' model obviously satisfies the requirement so understood, but it is not clear that it does on its more traditional understanding. Be that as it may, to build this condition into the *data* of the problem is at the very least contentious. Evans himself conceived of the bypass model as a *solution* to the problem of explaining self-knowledge in a way which could avoid the problem of positing the introspection of already given first-order mental states, rather than a constraint on any adequate account of self-knowledge. So, it is difficult to shake the impression that the methodology employed by Fernández tends, at least at times, to presuppose what it should be proved.

The third chapter shows how the bypass model can account for our knowledge of our own desires. Chapter four addresses the topic of Moore's paradox, in connection with the desideratum of the "assertiveness" of our self-ascriptions. That is to say, the idea that a self-ascription of a belief or a desire puts pressure on us to occupy the corresponding first-order mental states. Fernández claims that if the self-ascription is reached through bypass, it will be based on evidence (or lack thereof) which will make

it epistemically appropriate for us to endorse the corresponding first-order mental state. If that is the case, then Moore's paradox will represent a violation of this epistemic norm and will therefore be a case of "epistemic negligence" (cf. p. 127).

However, it is not clear how epistemic negligence can be enough to account for the irrationality of entertaining Moore-paradoxical contents, which is what Fernández himself wishes to explain. For, intuitively, irrationality requires more than mere negligence. Moreover, I think that a similar worry can be raised also with respect to Fernandez' discussion of self-deception in the last chapter of the book.

Fernández' *Transparent Minds* is a bold and instructive attempt to vindicate and develop Evans' claim that in making a self-ascription of belief, one's eyes are, so to speak, directed outward—upon the world. The resulting picture is not very convincing though. Perhaps this is a sign of the fact that Evans' intuition, as intriguing as it is, is not the key to the solution to the puzzle of self-knowledge.

Annalisa Coliva
University of Modena and Reggio Emilia
annalisa.coliva@unimore.it

HUBERTO MARRAUD. 2013. *¿Es lógic@? Análisis y evaluación de argumentos.* Madrid: Cátedra; LUIS VEGA REÑÓN. 2013. *La fauna de las falacias.* Madrid: Trotta; EDUARDO DE BUSTOS GUADAÑO. 2014. *Metáfora y argumentación: Teoría y práctica.* Madrid: Cátedra.

A poco más de cincuenta años de su nacimiento oficial,[1] la teoría de la argumentación constituye un campo de estudios inter- y trans-disciplinares cuyo éxito de implantación puede medirse por la proliferación de series internacionales de congresos,[2] revistas especializadas[3] y publicaciones de alto nivel y difusión mundial.[4]

En nuestro ámbito de habla española, el interés académico por el estudio de la argumentación también se está extendiendo en estos últimos años, sobre todo en medios filosóficos y lingüísticos, a medida que languidece, de manera cada vez más pa-

---

[1] Hace seis años se conmemoró el cincuentenario de la publicación de dos de los libros que suelen mencionarse como hitos que inauguraron el campo actual de la teoría de la argumentación: *The Uses of Argument* de Stephen Toulmin y *Traité de l'argumentation*, de Chaïm Perelman y Lucie Olbrechts-Tyteca, ambos publicados originalmente en 1958.

[2] Ya van diez ediciones del Congreso de la Ontario Society for the Study of Argumentation, ocho del de la International Society for the Study of Argumentation, cuatro de la serie Rhetoric in Society y en 2015 se inicia con renovado impulso y un evidente recambio generacional la serie European Conference on Argumentation (http://ecargument.org/).

[3] Las más conocidas y con mayor trayectoria del campo serían *Argumentation*, *Informal Logic* y *Argumentation and advocacy* a las que hay que añadir las más recientemente fundadas: *Argumentation et analyse du discours*, *Argumentation in context* o nuestra *Revista iberoamericana de argumentación*.

[4] Destacan, sobre todo, las series sobre argumentación y análisis del discurso de las prestigiosas editoriales Springer y John Benjamin's.