

## Book Reviews

Galeotti, Anna Elisabetta, *Political Self-Deception*.  
Cambridge: Cambridge University Press, 2018, pp. 261.

*Political Self-Deception* by Anna Elisabetta Galeotti is a brilliant example of how crucial the category of self-deception is in our attempts to correctly describe, conceptualize, and explain a wide variety of predicaments that go well beyond the domain of our private lives. Most of the literature on self-deception has thus far provided examples of the phenomenon when it surfaces in more or less troublesome events involving affective, emotional, and self-esteem related issues. Galeotti instead rightly takes self-deception to *also* have an impassable relevance for national and international political life: our tendency to self-deceiving is so pervasive in human psychology that nobody is immune to it, including political agents in the hands of whom is often the destiny of nations, the development of international relations, and even the future of our planet.

Galeotti's view of self-deception is framed by a widely defended paradigm of self-deception, long known as *motivationalism*: individuals are prone to self-deception on the basis of a more or less significant motivational set which can lead them to believe what a dispassionate analysis of the available evidence, and/or a dispassionate search for easily accessible evidence, would instead show to be false, or at least unlikely. Far from conducting such a dispassionate analysis of evidence and/or search for it, those agents do not operate an epistemically optimal treatment of evidence. Rather, they embrace false beliefs owing to the biasing effect of motivation over their rational cognition, so that they end up believing what they are independently interested in believing. Contrary evidence is thus avoided, discarded, misinterpreted, depending on the specific case. Interestingly, however, Galeotti's largely motivational theory is supplemented by at least two specifications:

- (1) Motivation may not be *per se* decisive to leading to self-deception, unless suitably favorable circumstances are also in place (35);
- (2) Defenders of motivationalism should be wary of embracing any "rampant" anti-agency view, amounting to a purely causal picture of self-deception, where an epistemic and practical agent proper is allegedly lacking—such a purely causal picture would also lead to the unpalatable consequence of making it difficult to attribute any sort of epistemic and practical responsibility to the agent who embarks on the self-deceptive maneuver (38).

It is important to emphasize why Galeotti thinks that (1) and (2) are in order. As to (1), by adding favorable circumstances to the motivational set of self-deception, we would be in a better position to try to overcome a troublesome objection, long known as the "selectivity problem". The "selectivity problem" objection roughly runs as follows:<sup>1</sup> motivation may not be sufficient to cause self-deception, as there may be people who do not end up self-deceiving, even when they are in the grip of a quite strong motivation to believe what they favor. If this is the case, as it seems it is, then self-deception is more selective than the mere presence of motivational factors predicts. In a word, we need additional conditions which can explain why motivation can casually go all the way down

<sup>1</sup> Bermúdez, J.L. 2000, "Self-Deception, Intentions and Contradictory Beliefs", *Analysis*, 60, 309-19.

to self-deception, when it does, and to diagnose what else is missing or present in the causal set of a subject when motivation is present, although self-deception does not happen. Thus, Galeotti argues (49) that motivation to believe what one favors is casually effective when the subject is *also* involved in a situation where there is a strong pressure to conclude that *p* is the case. Without such pressure, motivation might well remain inert.

As to (2), Galeotti is interested in counterbalancing any purely causal pictures of self-deception, because such views tend to conceive of the phenomenon as largely (if not entirely) passive, something that merely happens to us, and for which we couldn't be reasonably credited with any significant form of responsibility (38-44). By reintroducing the epistemic and practical operations of a subject that is not an entirely passive victim of motivational events, the subject is now best seen as practically and epistemically active. In this way responsibility for self-deception gets back into the picture. It is to be clearly emphasized that Galeotti is not endorsing any crudely *intentionalist* views of self-deception, that is, views where agents act *under the explicit and aware description of bringing it about their own self-deception, willingly, consciously and intentionally*. For there is much that agents do *not* see about themselves—above all, they do not see the hidden, distorting working of motivation over their cognition, and also the causal effect of the pressure to which they are exposed.<sup>2</sup> However, Galeotti is interested—and rightly so—in stressing that a subject is at work, and acts as an intentional agent does. For the subject assesses evidence, evaluates epistemic principles at hands, avoids or looks for new evidence, and so on. While failing to see that all of these epistemic endeavors are misled by the motivation towards a desired conclusion, and while also operating under an explicit description that does not include any attempts to deceive willingly him- or herself, the epistemic acts and the practical steps that he or she takes or avoids are nonetheless intentionally conducted. In other words, the subject works under the explicit description of wanting to take a certain step, assessing a certain piece of evidence, choosing to avoid another piece, and so forth.

I have already addressed<sup>3</sup> one first critical aspect regarding (2), and I now wish to cover one more point about (2). I start from discussing (2), and then I move on to an evaluation of (1).

I have argued that it is unlikely that classic motivationalism, as put forward by Al Mele,<sup>4</sup> for instance, can invariantly contain a genuine risk of being an anti-agency view, where the term “anti-agency” is interpreted as denying the presence in self-deception of an epistemic and practical agent proper. My sense is that such conflation may be due to a more or less hidden fallacy of equivocation bearing on the term “agency” as it appears in the “anti-agency” phrase, other than being a truly genuine issue. Arguably, when Mele put forwards what he dubbed as his “anti-agency” view of self-deception, he made use of the phrase “anti-agency” in order to mark a substantial difference from Davidson's inten-

<sup>2</sup> Scott-Kakures, D. 2002, “At ‘Permanent Risk’: Reasoning and Self-Knowledge in Self-Deception”, *Philosophy and Phenomenological Research*, LXV, 3, 576-603.

<sup>3</sup> Pedrini, P. 2020, “Autoinganno: con o senza agente? Sui rischi della fallacia di equivocatione per la spiegazione causale”, *Notizie di Politeia*, XXVI, 137, 194-96.

<sup>4</sup> Mele, A. 2001, *Self-Deception Unmasked*, Princeton: Princeton University Press, 13.

tionalism.<sup>5</sup> According to Davidson, self-deception is the product of an agent who willingly and intentionally brings it about his or her own self-deception. By attacking Davidson's intentionalism, Mele made use of the rhetoric of anti-intentionalism and anti-agency in order to emphasize that on his view of self-deception the subject is not someone who acts under the explicit and aware description of wanting to self-deceive. However, this does not imply that Mele leaves out of the picture any forms of intentional acts pursued by the agent under different descriptions. This seems to be confirmed by the conditions that Mele judges to be jointly sufficient to fall prey to self-deception.<sup>6</sup> They include a number of biased epistemic maneuvers that it is hard to consider as entirely passive. In all, if this reconstruction is correct, then any attempts to conflate classic motivationalism, even when rhetorically qualified as an "anti-agency" view, with the denial of any forms of agency, is perhaps due to a fallacy of equivocation on the term "agency" that might have somehow secretly operated over time. While in the heydays of Mele's anti-intentionalism was perhaps more obvious that the agency against which Mele's view was directed was the agency of someone acting under the explicit and aware description of causing oneself to be deceived, in Galeotti's theory the agency against which anti-intentionalism is directed is taken to risk to become a purely causal picture of self-deception, where the agent disappears. I am sympathetic with any views that emphasize the epistemic endeavors of agents when they self-deceive,<sup>7</sup> and I also agree that Mele may not have done the best interest of his theory when he used the phrase "anti-agency". Furthermore, it is a truly pointed issue to establish under what description then the agent acts, if he or she does not act under any explicit and aware description of bringing it about his or her self-deception. Nonetheless, I believe that it may be wrong to take the label "anti-agency view" to be invariantly equated to a purely causal account with no agent in place at all.

One second aspect regarding (2) that I briefly wish to touch upon has to do with the responsibility that attaches to a self-deceiver who actively works out his or her own self-deception, even if not with the explicit and aware view of causing his or her own self-deception. As said, Galeotti argues that self-deceivers intentionally enter the biased reasoning that is conducive to self-deception, even if they are not aware that self-deception awaits them at the end of the process. That would suffice to attribute to them a certain accountability for what they do. It is less clear, however, how this view works in details. For one thing, the prospect of this view dramatically changes if we start looking at it under the angle of the motivation that puts in motion the whole machine of the biased reasoning. If it turned out to be the case that self-deceivers are in fact *caused to reason* as they do by the motivation that alters their belief-formation cognitive process, would we be willing to attribute to them the responsibility for such reasoning? In other words, if we move back the focus of the theory from the form of their altered, intentional reasoning as such to the *cause* of such reasoning, we immediately seem to have fewer reasons to take them to be accountable for such reasoning. This is a notorious line of attack against all of those views which take awareness and intentionality of first-order acts and mental states to be a suffi-

<sup>5</sup> Davidson, D. 1985, "Deception and Division", in LePore, E. and McLaughlin, B. (eds.), *Actions and Events*, Oxford: Basil Blackwell.

<sup>6</sup> Mele, A. 2001, *Self-Deception Unmasked*, Princeton: Princeton University Press, 50-51.

<sup>7</sup> Pedrini, P. 2013, *L'autoinganno. Che cos'è e come funziona*, Roma-Bari: Laterza.

cient ground of responsibility. For causing factors can end up making *unaccountable* even fully aware and intentional first-order acts, if these causing factors are something on which the agent has ultimately no *control* or no other margins for correction and/or counteraction.

More demanding theorists may even require that we establish whether causing factors create conditions that ultimately leave the agent unfree to reason otherwise. I set all of these much thrown-around questions on responsibility *qua* connected to control and/or freedom because, even if Galeotti argues that she is going beyond the control model (64), I am left with the sense that her basis for attributing responsibility to self-deceivers relies on *much more* than the *mere* first-order intentional epistemic and practical operations of the subject who misleadingly and prejudicially treats evidence. As far as I can see, the issue can begin to be settled by analyzing *what other faculties* Galeotti is in fact attaching to first-order intentional epistemic and practical operations.

For the sake of pumping intuitions, let us consider two scenarios. The first scenario has us to imagine a subject who is unaware of the very existence of self-deception, of its working, and symptoms. At some point, this subject goes through the self-deceptive process. He or she might well be reasoning intentionally and consciously, as Galeotti diagnoses. However, by hypothesis, and for reasons to be explained, he or she has *no tools to detect* the phenomenon that is affecting and distorting his or her reasoning. In this case, it is hard to attribute full-blown responsibility to this agent.

Another scenario has instead us to imagine a subject who is equally unaware of the distorting effect of motivation over his or her cognition; however, the theorist here may point out that this subject is *in principle and as a matter of fact under certain conditions capable* of detecting that something is going wrong in his or her reasoning. For being this reasoning conscious and intentional, the subject is in the position to reflect upon such form of reasoning, to compare it with other instances of reasoning, and possibly also to make an inference from an aware, or easily discoverable, motivation to its causal, distorting role over cognition.

There are passages in the book which seem to confirm that Galeotti might have a preference for this second scenario (68). She seems to attach to the first-order intentional operation a variety of other capabilities that the subject either directly enjoys, or can at least retrospectively learn (68). One might wonder, however, whether this second scenario can constitute a sound normative, as well as empirical, basis for attributing responsibility. At the very least, this view requires, and empirically predicts, a good amount of (*ex ante* or *ex post*) surveillance over one's mental states and processes that many would find excessively demanding. Be that as it may, if Galeotti's view ultimately relies on the practice and/or the possible activation of these other faculties, then she might turn out to endorse a view where responsibility for self-deception is not merely based on the intentional first-order acts. Rather, her view seems to be ultimately based on some form of *control*, or *possible activation of control*, over one's mental states and processes.

Let me now turn back to (1). I have reasons to believe that (1) can hardly be a tenable solution to the "selectivity problem" as a problem of causation of self-deception. I can only offer here an outline of an argument against the alleged success of the strategy Galeotti adopts when she explains how favorable circumstance can be decisive to leading to self-deception. Other than blocking the selec-

tivity problem, favorable circumstances remain exposed to the reiteration of the selectivity problem against the causal compound made of motivation *plus* favorable conditions. That is to say, there may be people who do not self-deceive, even if they have the motivation and are *also* put in favorable circumstances. It may be true that the motivation to believe that *p*, coupled with circumstances of hard pressure, can increase the probability that one ends up self-deceiving. This may happen because pressure can erode the possible resistance that the subject, under less pressure, might still exert over the motivational thrust. If this is correct, then there must be something causally relevant in pressure, if a smaller number of motivated people resist against pressure. Yet, the addition of pressure as a means to making motivation stronger can't hardly be the end of the causal story, given that selectivity is still looming. Resistance to motivation and pressure thus seems to depend more on the psychological structure of the subject than on any other circumstances. Simply put, some people have a psychological structure that make them more resistant than others to various level of motivation and pressure. Even if we include the psychological structure of the subject in our causal analysis, it may be hard to adjudicate the question whether the psychology of the subject is causally decisive. Is the kind of psychological one has that is the *ultimate* determinant of self-deception? Or is it the vector computed by combining motivation and pressure with the psychological structure of the subject? This is a metaphysical topic for another, wider piece. But I think it is important to emphasize that there is a genuine issue here, which is urgent to adjudicate.

In all, this is a beautifully written, and excellently argued, book that by all means should become a must-read for a wide audience, including (although not limited to) students and scholars interested in political philosophy, international relations, social and political sciences, philosophy of mind and psychology, ethics, as well as all those who are active in policy making.

*University of Siena*

PATRIZIA PEDRINI

McCarroll, Christopher J., *Remembering from the Outside: Personal Memory and the Perspectival Mind*.

New York: Oxford University Press, 2018, pp. xx + 220.

The theme of observer memory might strike one who hasn't read McCarroll's book as somewhat narrow or specialized. Anyone who has read the book, however, will understand that it in fact intersects with a wide range of issues in the philosophy of memory and beyond. The book is thoroughly researched, rigorously argued, and might be read with profit by philosophers working not only on memory but also on perception, imagination, and language. It might also profitably be read by psychologists working on any of these topics, for McCarroll both displays an impressive mastery of the relevant empirical literature and makes use of philosophical tools to shed considerable new light on the conceptual puzzles to which it gives rise.

In field perspective memory (FPM), one remembers an event from the point of view from which one originally experienced it; in observer perspective memory (OPM), in contrast, one remembers an event from a point of view other than that from which one originally experienced it, seeing oneself in the remem-

bered scene. The central question of the book is whether OPMs can be fully genuine or faithful to the past. McCarroll claims—in opposition to a claim found throughout the philosophical and psychological literatures—that it can be, and this in a very precise sense. We can distinguish between the truth of a retrieved memory and its authenticity (Bernecker 2008):<sup>1</sup> a memory is true if it is accurate with respect to the remembered event; it is authentic if it is accurate with respect to the subject’s experience of the remembered event. What McCarroll claims is that OPMs can be both true and authentic.

The claim that OPMs can be true is unproblematic: one’s memory of an event might clearly be accurate with respect to the event even if the point of view from which one remembers it differs from the point of view from which one experienced it. It is the claim—and McCarroll makes it clear that he does indeed mean to defend this claim—that OPMs can be authentic that is surprising. How, given that one did not see oneself while experiencing the event, might a memory in which one sees oneself be accurate with respect to one’s experience of the event? (Compare two photos, taken from different angles, of the same scene: they might both match the scene, but they will not match each other.) After reviewing the literature on OPM in chapter 1, McCarroll develops, in chapters 2 and 3, a two-part framework designed in part to answer this question. The framework combines a “constructive encoding” approach and a “reconstructive retrieval” approach. The basic idea behind the former is that constructive processes occurring during encoding may shape the content of a stored memory. The basic idea behind the latter is that reconstructive processes occurring during retrieval may shape the content of a retrieved memory. Both of these ideas are empirically well-established, but the way in which McCarroll builds on them to defend the claim that OPMs can be authentic is highly original.

According to the reconstructive retrieval approach, the new content that figures in an OPM—including a visual representation of the rememberer himself—is sometimes the product of reconstruction at retrieval. This approach thus does not by itself imply that OPMs can be authentic. According to the constructive encoding approach, the apparently new content that figures in an OPM is sometimes the product of construction at encoding; the apparently new content may, in other words, not be new at all. This approach thus implies that OPMs can be authentic: the problematic components of the content of an OPM, including, in particular, the visual representation of the rememberer himself, may, in some cases, have figured in the corresponding earlier experiences. In short, there is an important sense in which one does sometimes see oneself while experiencing an event, a sense in which one sometimes has “observer perspective experiences” (OPEs).

The notion of an OPE is both the most provocative and the most problematic element of McCarroll’s book. He emphasizes that he is not interested in what we might call literal OPEs, experiences in which one literally entertains a visual representation of oneself while experiencing an event (e.g., by visually imagining oneself from a hypothetical observer’s point of view). Thus his claim is not that an OPM might be authentic because the apparently new content that figures in it figured, in the same, visual form, in the corresponding OPE. It is, instead, that an OPM might be authentic because the apparently new content that figures in it figured, in a different, nonvisual form, in the OPE. This entails

<sup>1</sup> Bernecker, S. 2008, *The Metaphysics of Memory*, Netherlands: Springer.

that the content of a nonliteral OPE can be equivalent to the content of a literal OPM, and McCarroll goes to considerable effort to show that this is the case, arguing that experience has a multimodal character and that information in one modality might be “translated” into another. When giving a public talk, for example, one experiences the scene from one’s own visual point of view, but one might (say, if one is feeling self-conscious) also experience it from an emotional observer point of view; this emotional content can then be translated, during encoding, into a visual representation of one’s self.

McCarroll’s strategy here is ingenuous. It does, however, face two obvious problems. First, the notion of an OPE is highly speculative. McCarroll borrows the notion from Nigro and Neisser’s foundational (1983) paper.<sup>2</sup> Nigro and Neisser do not, however, provide any real evidence for the occurrence of OPEs, and the concept of an OPE has played no role in subsequent research on OPM. McCarroll does point to evidence from a variety of sources suggesting that the self may be present in experience in a variety of (nonvisual) ways, but this evidence falls short of indicating that ordinary experiences may include content of the required sort. Second, the notion of translation is likewise highly speculative. The claim about authenticity presupposes not only that a nonvisual representation of the self may, via the translation process, give rise to a visual representation of the self, but also that it may do so without generating new content, for, if a retrieved memory includes content that was not included in the experience, it is by definition inauthentic. McCarroll does point to evidence suggesting that information in one modality can be translated into another modality; this evidence does not, however, indicate that a nonvisual representation can be translated into a visual representation without the addition of new content.

In order to surmount these problems, McCarroll might relax the standard of authenticity so as to allow a memory to be authentic as long as it includes at most a moderate amount of new content. Once the standard is relaxed, however, it becomes hard to see why we should care about it at all: if genuine memory is compatible with the addition of a moderate amount of new content, why think that it is incompatible with the addition of a large amount of new content? Two further moves suggest themselves. First, McCarroll might weaken the concept of OPM: if the OPMs in which he is interested are not literal OPMs—i.e., if they do not include visual representations of the self but only, say, emotional representations of the self—the authenticity claim becomes much more plausible. Second, he might strengthen the concept of OPE: if the OPEs in which he is interested are literal OPEs—i.e., if they do include visual representations of the self—then the authenticity claim again becomes much more plausible. The cost of making either of these moves is, however, a significant reduction in the interest of the authenticity claim. The claim that literal OPMs can be authentic with respect to literal OPEs and the claim that nonliteral OPMs can be authentic with respect to nonliteral OPEs are much less surprising than the claim that literal OPMs can be authentic with respect to nonliteral OPEs.

There may, of course, be further moves that are open to McCarroll. The intention of these remarks is not to show that his argument does not succeed but simply to show that there are ways of pushing back against it. And if it should eventually turn out that the argument does not succeed, McCarroll will never-

<sup>2</sup> Nigro, G. and Neisser, U. 1983, “Point of View in Personal Memories”, *Cognitive Psychology*, 15, 4, 467-82.

theless have enabled us to learn a great deal about the nature and limits of memory's faithfulness to the past; this alone is sufficient to ensure that his book will stand as a major contribution to the philosophy of memory.

McCarroll's argument for the authenticity claim will be of interest not only to philosophers of memory but also to philosophers of perception. Subsequent chapters will be of interest to philosophers working on topics including imagination and language. Chapters 4 and 5 engage critically with Vendler's (1979)<sup>3</sup> claim that imagining "from the outside" is just a special case of imagining "from the inside", arguing that Vendler goes wrong by overlooking the possibility of an unoccupied point of view in visual imagery. Chapter 6 engages critically with François Recanati's work, arguing that Recanati's (2007) framework<sup>4</sup> can be modified so as to accommodate a form of implicit *de se* thought that is both first-personal and from-the-outside. Chapter 7 of the book brings things to a conclusion by drawing together the various threads of the overall argument.

The quantity of published philosophical work on memory has increased rapidly over the past few years. Much of this work is of the highest quality. Even against this background, however, McCarroll's book stands out as one of the most important contributions to the area in many years. The publication of the book is a major event in the philosophy of memory. I look forward to engaging with it in my own future work, and I have no doubt that many others will as well.<sup>5</sup>

*Centre for Philosophy of Memory*  
*Université Grenoble Alpes*

KOURKEN MICHAELIAN

<sup>3</sup> Vendler, Z. 1979, "Vicarious Experience", *Revue de Métaphysique et de Morale*, 84, 2, 161-73.

<sup>4</sup> Recanati, F. 2007, *Perspectival Thought: A Plea for (Moderate) Relativism*, Oxford: Oxford University Press.

<sup>5</sup> McCarroll is currently employed as a postdoc at the Centre for Philosophy of Memory, which I direct. I had read the book and agreed to write this review well before he was recruited. This work is supported by the French National Research Agency in the framework of the "Investissements d'avenir" program (ANR-15-IDEX-02).

Mendelovici, Angela, *The Phenomenal Basis of Intentionality*.  
Oxford: Oxford University Press, 2018, pp. xviii + 275.

What is the deep nature of intentionality? What is its source? What relation is there between intentionality and phenomenal consciousness? These are some of the main questions addressed by Angela Mendelovici in her recent, well written and original book *The Phenomenal Basis of Intentionality*. Mendelovici's proposal, as the title of her book clearly suggests, belongs to that family of theories that take a "consciousness-first" approach to intentionality. Such an approach is endorsed by all the advocates of the so called "Phenomenal Intentionality Theory" (PIT) (people like Loar, Searle, Siewert, Strawson, Kriegel, Horgan, Tienson, Pitt, Farkas, Chalmers, Smithies, Montague, to mention just a few of them). For PIT's friends, intentionality has an experiential-phenomenal nature and has its source in phenomenal consciousness. According to Mendelovici, this is so because intentionality is phenomenal consciousness, in the sense that intentionali-

ty is not merely *a species* of phenomenal consciousness (as most of PIT's advocates claim), but rather *its only species* and therefore truly identical with it. This peculiar way of accounting for the relationship between these two (at least conceptually distinct) properties qualifies her proposal as a strong identity version of PIT, according to which phenomenal intentionality is the only kind of intentionality (there is no non-phenomenal kind of intentionality) and it is identical to phenomenal consciousness. Other versions of PIT provide different accounts of the relationship between intentionality and consciousness, either in terms of grounding, or in terms of constitution, or of realization. According to Mendelovici, strong identity PIT provides the clearest possible answer to the question regarding what kind of property intentionality is. By contrast, she claims, any non-identity account leaves the nature of intentionality ultimately unexplained. Whether she is right or wrong in so claiming, her book is definitely worth-reading, if only for the fact that it provides the most articulated and sophisticated presentation of an original theoretical proposal in the philosophical debate on the nature of intentionality.

Before presenting my critical remarks, let me provide a sketchy presentation of the book's overall structure and main claims. The book, which consists of six parts, can be taken to be organized into four main topic sections. In the first one Mendelovici presents the methodology that she adopts. This section deals with the following question: How can we fix reference on intentionality? In her view, the traditional answer according to which intentionality is the property that mental states originally possess of being about, being directed towards something (a property, a thing, a state of affairs) is too vague to constitute the starting point of the inquiry. Even though the notions of *aboutness* and *directedness* gesture towards the target property, they are, in her view, too fuzzy to provide a firm grip on it. In order to remedy such a lack of precision, she suggests to make use of an ostensive reference-fixing definition which characterizes intentionality as "that feature, whatever it is, that we at least sometimes notice in ourselves and are tempted to describe using representational terms like 'aboutness' and 'directedness'" (5). She criticizes other ways of fixing reference on intentionality (ways that make reference to some alleged roles that intentionality is supposed to play) because they do not individuate, in her view, intentionality, but different, albeit related, notions. This section is followed by a *pars destruens* whose aim is to assess whether PIT's main competitors (namely: tracking theories and functional role theories both in their short and in their long-arm version) are true and empirically adequate. In order to show their empirical inadequacy, she makes use of what she labels the "mismatch problem" that she illustrates by presenting cases in which the contents attributed by the competing theories do not match the contents we have "theory-independent reasons" to think we represent. In her view, the competing theories are guilty of two kinds of errors: an "error of commission" (in so far as they include in the representation's content material that is not represented) and an "error of omission" (in so far as they do not include material that is represented). As to the faultiness of the competing theories, she presents what she labels the "Real Problem" according to which what is ultimately wrong with those theories is not so much that their alleged content-conferring external relations sometimes "grab" the wrong content, but rather that they cannot "grab" anything at all, because genuine intentionality cannot arise from mere external relations with the world. What would then be the right source of genuine intentionality according to Mendelovici? This is the question

that she addresses in the third topic section of her book, in which she presents her positive proposal. According to it, intentionality cannot arise from anything different from phenomenal consciousness because only the latter has the right ingredient to give rise to intentionality. This is so, in her view, because intentionality just is phenomenal consciousness (that is precisely what the identity account that she endorses amounts to). She then concludes by showing that her theory is also not clearly empirically inadequate in so far as it can accommodate all cases of intentionality, including those that are commonly thought to pose problems for PIT (these cases include thoughts, non-conscious mental states either standing or occurrent, states representing rich descriptive contents or object-involving contents or more generally broad contents).

In my critical discussion of Mendelovici's proposal I shall mainly focus on two issues. The first one has to do with the methodology that she adopts and recommends. My main question here is the following: Is she right in claiming that her suggested way of fixing reference on intentionality is theory-independent? This point is of the outmost importance in order to assess whether the starting point of the inquiry is neutral enough as not to compromise from the very beginning its entire development. The second issue has to do with whether her phenomenally-based account of intentionality is ultimately able to provide a satisfactory answer to what, according to many people, is *the* problem that any theory of intentionality worthy of its name has to address, namely to account for how our mind succeeds in "getting in touch" with the world. The discussion of this point is preceded by an assessment of the criticisms she raises against PIT's rival theories. Let me now address these issues starting from the first one.

To begin with, it is important to stress that the ostensive definition that she introduces to fix reference on the property of intentionality does not ultimately dispense with the traditional characterization of it in terms of "aboutness" and "directedness". Rather, it incorporates such a characterization together with the further condition according to which intentionality is a feature that we at least sometimes (i.e. in what she considers "paradigm cases" such as perceptual states, but also thoughts and judgments) notice in ourselves. She tries to minimize the role that the traditional characterization plays in her picture by claiming that the ostensive definition only *mentions* the aforementioned representational terms without using them. Actually, the role they play seems to be more substantive than that in so far as it is only by making use of them that it is possible to unify under a single label the huge variety of things we can introspectively notice, no matter how different they (phenomenally) appear from each other (compare a perceptual state with a judgment, for example). This said, let us consider the other part of her suggested definition. As I have said, the definition of the subject matter of an inquiry should not be committed to any controversial way of conceiving it. Does Mendelovici's definition satisfy this requirement? Well, as a matter of fact, in order for something to be introspectively noticeable it has to have an experiential-phenomenal nature and that intentionality does actually possess such a nature is a point that not everyone in the philosophical arena is ready to accept. Think for example of theorists like Fodor, Dretske and Millikan for whom intentionality is a property that can be studied by natural sciences by making use of their investigation methodologies. Any such theorist would find Mendelovici's starting point question-begging and theory-laden. It has to be stressed that, according to Mendelovici, what introspection reveals is not so much the *nature* of intentionality, but rather, as she says, its *superficial*

*character*. What she means by this somewhat vague expression is that introspection gives us knowledge of which contents our conscious occurrent mental states have. I want to make two critical remarks on this point. First. Even granting that we do have some sort of introspective access to the mental contents of (at least some of) our conscious occurrent mental states (but, as we will see in a moment, this is an issue that should not be taken for granted), it does not follow that we also have introspective access to intentionality, that is to *the feature* of our mental states *in virtue of which* they have the contents they have and a content at all. Mendelovici does not seem to be aware of this problem, for she says: “When we introspectively notice intentionality, we do so at least by introspectively noticing our contents. Indeed, it might be that there is nothing more to notice when we notice intentionality than those intentional contents” (note 7, p. 8). In any case, and this is my second remark, even assuming that the introspective noticeability of contents suffices for the introspective noticeability of intentionality, the fact remains that such a claim about mental contents presupposes an internalist metaphysical picture of their nature and individuating conditions that is anything but uncontroversial and non-committal. Mendelovici is well aware of the fact that if externalism were true, then her suggested methodology for finding out about our intentional states and their contents would prove ineffective. She discusses this point by observing that it is generally taken to be an objection to externalism that its adoption makes it difficult, if not impossible, to account for the introspective accessibility of mental contents, which is why most advocates of externalism try to defend the claim that their metaphysical picture of the mind is compatible with a plausible account of self-knowledge. Her conclusion on this point is quite hasty: “If this attitude is correct [the so called “compatibilist attitude” endorsed by most externalists], then if there is an incompatibility between the assumption underlying my methods and externalism, it is externalism that should be rejected” (66).

Let me recap the main points of my critical assessment of the methodology that she adopts: a) the reference-fixing definition that she introduces seems to be no less fuzzy than the traditional one, in so far as it not only incorporates, but makes substantive use of it; b) the claim that intentionality is introspectively noticeable is theory-laden, since it presupposes that intentionality is an experiential property; c) it is disputable that the introspective noticeability of content suffices for the introspective noticeability of intentionality; d) the claim that mental content is introspectively noticeable presupposes a non-neutral picture of mental content. In my view, these points provide strong, albeit non-conclusive, evidence in support of the claim that the starting point of Mendelovici’s inquiry is not as theoretically neutral as it should be. This seems to me to be enough to put pressure on her claim that her recommended methodology is theory-independent.

The issue of the theory-independency is relevant also for the second issue I want to address. As I said at the beginning, the non-neutrality of the starting point of an inquiry very often propagates to the inquiry itself. In my view, this is precisely what happens in Mendelovici’s case, in particular as regards the kind of arguments that she presents in support of the claim that PIT’s rival theories are both empirically inadequate and wrong because they do not have the right ingredients to account for how our minds succeed in getting in touch with the world. My main aim here is to assess, first, whether those arguments go through without assuming any controversial and theory-laden assumption and, second, whether strong identity PIT is ultimately able to properly account for that issue.

Let us start from the first point. As anticipated in the presentation of the book, Mendelovici ascribes to the competing theories two kinds of error that should falsify the theories' predictions as to what a given mental state would or would not represent given the truth of those theories. The arguments she presents in support of her criticism have the following structure: "If theory T is true, then the representation R has the content C / Representation R does not have the content C / Therefore..." (as regards the error of commission) and "If theory T is true, then the representation R does not have the content C / Representation R does have the content C / Therefore..." (as regards the error of omission). The problem I here see concerns the two arguments' second premises. Do we have any theory-independent way of finding out which content a given mental state has or does not have? Mendelovici thinks we do: introspection and considerations of psychological role are precisely two such ways. Well, are they truly theory-independent as she claims? Let us consider the following point. As a moment's reflection shows, neither argument could go through unless an implicit premise were in place, respectively: (i) a given representation cannot represent a content unless it is a content the subject "feels" she is entertaining (as for the argument in support of the error of commission) and (ii) a representation cannot avoid representing a given content if it is one the subject "feels" she is entertaining (as for the argument in support of the error of omission). These two further premises hint at what she labels the "psychological involvement" of mental contents, which ultimately amounts to the idea that which contents we entertain is wholly transparent to us, because we are the authors of those very contents. As before, this claim commits itself to a radically internalist picture and such a commitment, which is licensed by her suggested ways of finding out which contents our mental states have, is anything but theory-independent. Or so I claim.

The issue concerning the psychological involvement of mental content comes up again in the "Real problem" that she raises against the competing theories: "theories relying on tracking relations not only fail to attribute psychologically involved content in mismatch cases but, worse, preclude the contents they attribute from being psychologically involved in the first place" (79). According to Mendelovici, what ultimately shows that the competing theories are wrong is their inability to account for how our minds can make psychological contact with the external world. As she says: "The Real Problem with the relation view is that it's hard to see how any relation to distinctly existing items can make them entertained or otherwise intentionally represented [...] No ordinary relation can allow us to literally entertain tables and chairs, to take hold of objects existing in the concrete world and bring them into our minds to make them available to our cognitive system" (204). I think that the way in which she frames the problem she is confronting with is strictly conditioned by her internalist picture of the mind. By assuming such a picture, the problem becomes that of accounting for how the gulf between mind and world can be bridged, so as to bring inside the mind what is outside of it. According to Mendelovici, no non-phenomenal theory of intentionality is able to solve this problem. Well, perhaps she is right (maybe because, once so framed, the problem becomes intractable). But is her phenomenal intentionality theory able to account for the psychological involvement of the world after all? About this point I find the following passage very enlightening: "concretely existing objects, properties, and states of affairs exist distinctly and usually independently of us, our cognitive systems, and our intentional states. So, it's hard to see how any relation we can

bear to these distinctly and independently existing items can make them psychologically involved. So contents can't be such items. *Tables, chairs, and obtaining states of affairs are not thinkables, experienceables, entertainables, or, more generally, (intentionally) representables. They are not the kinds of things that can play the role of intentional contents*" (205, emphasis mine). Well, is this a solution to the problem raised or just a capitulation? And, in the latter case, can a theory of intentionality worthy of its name be compatible with such a capitulation? In my view, even granting that Mendelovici is right in claiming that tracking and functional role theories of intentionality leave unexplained an issue that any adequate theory of intentionality ought to address, the fact remains that, on balance, not even her account seems able to explain how the world, the real world made of concretely existing objects like you and me and tables and chairs, can be psychologically involved.

It ultimately turns out that her account of intentionality is highly revisionist. If she is right, intentionality does not possess most of the properties that we tend to ascribed to it, such as: being relatively abundant, externalistic, relational. All this can be accepted. But what about the trait according to which intentionality is the directedness of our mental states at something *beyond* themselves (ordinarily, concrete objects and properties) which could exist independently of being experienced? Is this a dispensable trait or is it rather a constitutive element of the very nature of intentionality, as most people (even within the PIT's camp) claim? Personally, I am on the side of all those who take such a trait to be indispensable. I therefore believe that in so far as Mendelovici's picture gets rid of it, one can legitimately conclude that it gets rid of intentionality through and through. But if this is so, what is her theory ultimately a theory of? My suspicion is that her attempt at overcoming the fuzziness of the traditional characterization of intentionality by trying to pin down the property by means of mainly experientially-based criteria ends up meeting the same fate as other attempts already made in the past, like for example Chisholm's, namely that of individuating not intentionality, but a property that, albeit closely related to intentionality, does not ultimately coincide with it. My suspicion is that the property in question is the *presentationality* that accompanies any conscious occurrent intentional state, that is the property that accounts for the fact that consciously representing something is always accompanied by a presentation to the subject of what she is thereby representing. It is a pity that the distinction between representation and presentation is rarely thematised within the PIT's camp. I personally think that such a thematisation would highly improve the ongoing debate on intentionality.

It has to be said that Mendelovici shows great awareness of all the problems that her proposal raises. Not only that. Actually she addresses them and, more often than not, provides interesting and plausible solutions. Regardless of whether one finds them satisfactory or not, one must allow her an uncommon capacity of making a radical and provocative position look eminently reasonable. To conclude, let me say that Mendelovici's book is to be warmly welcomed: if it contains much to disagree with, it also contains much to learn.