IS SELF-DECEPTION PRETENSE?*

JOSÉ EDUARDO PORCHER

Programa de Pós-Graduação em Filosofia
Universidade Federal do Rio Grande do Sul
Av. Bento Gonçalves, 9500
91501-970 PORTO ALEGRE, RS
BRASIL

jeporcher@gmail.com

Abstract: I assess Tamar Gendler's (2007) account of self-deception according to which its characteristic state is not belief, but *imaginative pretense*. After giving an overview of the literature and presenting the conceptual puzzles engendered by the notion of self-deception, I introduce Gendler's account, which emerges as a rival to practically all extant accounts of self-deception. I object to it by first arguing that her argument for abandoning belief as the characteristic state of self-deception conflates the state of belief and the process of belief-formation when interpreting David Velleman's (2000) thesis that belief is an essentially truth-directed attitude. I then call attention to the fact that Velleman's argument for the identity of motivational role between belief and imagining, on which Gendler's argument for self-deception as pretense depends, conflates two senses of 'motivational role'—a stronger but implausible sense and a weaker but explanatorily irrelevant sense. Finally, I introduce Neil Van Leeuwen's (2009) argument to the effect that belief is the practical ground of all non-belief cognitive attitudes in circum-stances wherein the latter prompt action. I apply this framework to Gendler's account to ultimately show that imaginative pretense fails to explain the existence of voluntary actions which result from self-deception.

Keywords: Self-deception. Belief. Imagination. Pretense. Make-believe. The aim of belief. Truth-directedness. Motivation. Action.

---

1. Introduction

Self-deception is a psychological phenomenon with which every human being is familiar (with the exception, perhaps, of those who are too good at it). There is no shortage of examples, but maybe it will help if we confine ourselves to those which are relatable and less intricate.[1] Consider the following examples. A mother insists that her son, who went missing in action during the war and whose apparently lifeless body was sighted by a fellow soldier, is still alive. A wife refuses to accept that her husband's increasingly long absences and failure to explain these with any measure of credibility are indication of foul play. An oncologist ignores his own unmistakable symptoms of cancer, deflecting his family's attempts to discuss his condition and taking to writing letters to distant friends and relatives, intimating farewells, and even drafting a will.

Notwithstanding our familiarity with this type of situation, and although the folk-psychological concept of self-deception is used by us every day, we have yet to come up with a successful explanation of all its aspects. And while it may not be obvious why philosophers (as opposed to experimental psychologists, neuroscientists, and behavioral economists) should devote themselves to analyzing in depth the notion of self-deception, the simple explanation is that, while the experience is common and real enough, the concept of self-deception still eludes us.

The phenomenon itself has been portrayed since the earliest texts that have come down to us. The oldest allegory of self-deception I

---

[1] I was first drawn to the subject of self-deception through literature, especially Ibsen's *The Wild Duck* and O'Neill's *The Iceman Cometh*, and recommend these works as a source of beautifully rendered, ostensible self-deception. For analyses of these plays within investigations of self-deception, see Martin (1986, pp. 110-6) and Neu (2000).

know is in the Hebrew Bible. Part of the second Book of Samuel narrates the weaknesses and failures of David's kingship, two of which were adultery and murder. Upon falling for Bathsheba, the beautiful wife of a righteous soldier named Uriah, David sends him out to the front where the ongoing war is fiercest. Uriah subsequently dies in combat, fulfilling David's plan, and as soon as Bathsheba is done mourning her husband's death, David brings her to his house and she becomes his wife. Displeased with this, God sends Nathan, the court prophet, to reprimand David. He does this by telling David a story about a rich man who, instead of choosing among his very large number of sheep and cattle, takes a poor man's only lamb—a beloved animal that grew up with his children, shared his food, and drank from his own cup—to prepare a meal for a traveler who had come to him (2 Samuel 11-12). Upon hearing this, David is incensed and demands to know who the rich man is, proclaiming that he must pay for that lamb four times over and be sentenced to death. To what Nathan predictably replies that David himself is the man in the allegory. He finally sees the light.[2]

However, to my knowledge, Kant was the first author to recognize that there is something amiss with the very concept of self-deception. He formulated the now widely known puzzle attached to this concept in the second part of *The Metaphysics of Morals*: 'It is easy to show that man is actually guilty of many inner lies, but it seems more

---

[2] There are other ancient mentions of self-deception in such texts as Plato's *Cratylus* (428d) and Paul's *Epistle to the Galatians* (6:3), but Augustine was perhaps the first writer to actually develop the subject in the course of examining his own ways before conversion in the *Confessions*. He was the forerunner of a tradition of Christian writers who pursued the theme of self-ignorance—a tradition that continued with Pascal's *Pensées* down to many 17th and 18th century British thinkers such as Daniel Dyke, Richard Baxter, Joseph Butler, and John Mason. The subject eventually received a more secular, if cursory, treatment in Adam Smith's *Theory of Moral Sentiments*.

difficult to explain how they are possible; for a lie requires a second person whom one intends to deceive, whereas to deceive oneself on purpose seems to contain a contradiction' (1797/1996, p. 183). This puzzle, however, did not give rise to a specialized debate up until the translation of Sartre's *Being and Nothingness* into English. In his discussion of 'bad faith,' Sartre recognizes and elaborates on the same contradiction: 'I must know in my capacity as deceiver the truth which is hidden from me in my capacity as the one deceived. Better yet, I must know the truth very exactly *in order* to conceal it more carefully— and this not at two different moments, which at a pinch would allow us to reestablish a semblance of duality—but in the unitary structure of a single project' (1949/1957, p. 49).

Perhaps it is easier now to envisage why self-deception has held interest for philosophers (especially since Raphael Demos published 'Lying to oneself' in 1960). As David Pears succinctly put it, 'self-deception is an irritating concept. Its supposed denotation is far from clear and, if its connotation is taken literally, it cannot really have any denotation' (1984, p. 25). Which is to say that, apart from the very difficulty of arriving at a consensual definition, the very word 'self-deception' carries with it an air of impossibility if we take it to mean exactly what it seems to mean. As Alfred Mele (2001, p. 6) points out, on close inspection two puzzles arise from a literal interpretation, each of which is derived from one of two lexical assumptions:

> 1.   By definition, person A deceives person B into believing that p only if A knows, or at least believes, that $\neg p$ and causes B to believe that p.
> 2.   By definition, deceiving is an intentional activity (i.e., non-intentional deceiving is conceptually impossible).

The first puzzle, then, arises from the recognition that if I deceive myself in the same manner in which I deceive someone else, it seems that I am in an impossible state of mind, namely, that of believing two contradictory propositions p and $\neg p$ simultaneously. (This is, of course, not to say that self-deceivers believe a contradiction, but only that they have a pair of beliefs the content of which is logically incompatible.) The second puzzle, on the other hand, arises from the recognition that if I literally deceive myself, it seems that I engage in the impossible process of intentionally bringing myself to believe something that I myself believe to be false. Mele (2001) calls these problems the *static* and *dynamic* puzzles of self-deception, respectively.[3]

In response to the aura of paradox that results from a lexical interpretation of 'self-deception,' some philosophers such as Mary Haight (1980), David Kipp (1980), and Kenneth Gergen (1985) have become convinced of the impossibility of self-deception itself. And as much as we might think that the strategy of denying the existence of a commonly experienced phenomenon would have gone out with the days of literalistic ordinary language conceptual analysis, Steffen Borge has recently argued that 'there is no such thing as self-deception ... what

---

[3] In what follows I will focus on the first of these problems; after all, my intended scope is the import of the examination of borderline phenomena for the study of belief, and not a full account of the phenomena discussed. The study of the dynamic by which self-deception occurs inevitably engulfs one in action-theoretic debates about intentionality which I put aside for reasons of space and relevance. Suffice it to say that the intentional character of self-deception has been challenged, since unintentional deception, far from being conceptually impossible, is commonplace (see Barnes 1997). Lying, on the other hand, is intentional by definition, and the second lexical assumption mentioned above stems exactly from a confusion of 'deception' and 'lying'. See Carson (2009) for a thorough treatment of definitions of lying and deception (and the related concepts of withholding information, keeping someone in the dark, bullshit, spin, and half-truths).

has formerly been known as self-deception is rather a failure to understand, or lack of awareness of, one's emotional life and its influence on us' (2003, p. 1).

That being said, there has always been a fairly general consensus in the literature that self-deception *does* exist. One of the most famous strategies that have been undertaken to explain literal self-deception has been developed by David Pears (1984) and Donald Davidson (1985). Both their views rest on the Freudian idea that the best way to account for the phenomenon is to *split* the person. Pears, for instance, argues that there is a subsystem built around the self-deceptive belief and that it is organized like a person—a homunculus, one might say. 'Although it is a separate center of agency within the whole person,' Pears notes, 'it is, from its own point of view, entirely rational. It wants the main system to form the irrational belief and it is aware that it will not form it, if the cautionary belief is allowed to intervene. So with perfect rationality it stops its intervention' (1984, p. 87).[4]

It's not difficult to see how this would solve the pending difficulties. Pears converts the problematic characterization '*A* deceives *A*' that resulted from a lexical reading into the non-problematic '*A* deceives *B*,' where *A* and *B* are different subsystems of agency within a

---

[4] Note that by offering a psychological explanation to the conceptual puzzles of self-deception, Pears shifts the focus of investigation from an analysis of self-deception *in the abstract* (i.e., human or otherwise) to an analysis of human self-deception. Though self-deception might be multiply realizable, the literature on self-deception has abandoned purely conceptual investigation (as it was undertaken from the early 1960s to the mid-1980s) and is now increasingly informed by empirical research in cognitive science (see, e.g., Mele 2001, Gendler 2007, Davies 2009, and McKay et al. 2009). Thanks to an anonymous referee for bringing the issue of the task of an analysis of self-deception to my attention.

presumably unified system, namely, the person. Because the roles of deceiver and deceived are played by different centers of agency, the aura of paradox disappears. However, this sort of explanation faces its own difficulties. As Mark Johnston (1988, p. 64) observes, the homuncular explanation replaces a contradictory description of the self-deceiver with its own set of psychological puzzles: how can the deceiving subsystem have the capacities to perpetrate the deception? Why should the deceiving subsystem be interested in the deception? Does it suppose that it knows what it is best for the deceived system to believe?[5]

On the other hand, Davidson proposed what we may call a *functional* division, which bypasses the aforementioned charge to homuncular explanations. His view is that all that is needed is a boundary between conflicting attitudes—there would be no contradiction in believing contradictory propositions if they didn't come in contact with each other. Davidson claims that it is the drawing of such a boundary between our inconsistent beliefs which constitutes the irrational step involved in self-deception, and that this step is assisted by the non-observance of what Hempel and Carnap called the *requirement of total evidence for inductive reasoning*—a normative

---

[5] Because it assimilates the intrapersonal to the interpersonal case, this kind of explanation may also be seen to bring about problems concerning motivation, as Gendler (2007, p. 235) observes: 'If one of the subpersons (truly) believes that p and does not believe that $\neg p$, and if that subperson is bothered by this and wishes it were not the case, why would she find it psychologically fruitful intentionally to bring someone *else* to believe the opposite?'. However, as an anonymous referee has pointed out to me, Johnston's and Gendler's criticisms of the so-called homuncular approach may miss the mark by ignoring that the division of psychological labor between a conscious and an unconscious self has an evolutionary point. See, e.g., Evans and Frankish (2009) and Kahneman (2011).

principle that enjoins us to give credence to the hypothesis most highly supported by all available relevant evidence when choosing among a set of mutually exclusive hypotheses. Thus, Davidson formulates the conditions under which an agent A is self-deceived with respect to a proposition p in the following manner: 'A has evidence on the basis of which he believes that p is more apt to be true than its negation; the thought that p, or the thought that he ought rationally to believe p, motivates A to act in such a way as to cause himself to believe the negation of p' (1985, p. 88).

Despite the modest amount of division required to make sense of this proposal, most authors since Davidson have abandoned the literal reading altogether. Many have devised theories of self-deception in accordance with the spirit of Davidson's proposal, but they have chosen not to draw the unwanted consequence that the subject believes *both* propositions involved. For proponents of this view, self-deception is not to be understood as a reflexive form of deception, in the same manner that self-teaching is not understood as a reflexive form of teaching. The 'deception' in self-deception must be understood as a *metaphor* (an observation that originated as early as Canfield and Gustafson 1962). Such theorists are, however, left with the task of explaining what exactly the mental states involved in self-deception are, and how they are formed and maintained: in other words, the task of explaining what the metaphor stands for. In their efforts to make sense of non-literal self-deception, the main disagreement has been over what kind of attitude is the *product* of self-deception (i.e., the unwarranted proposition to which the self-deceiver is motivated to give assent), and also over whether self-deceived subjects retain a belief in what Neil Van Leeuwen (2007) calls the *doxastic alternative* (i.e., the warranted proposition the content from which the self-deceiver is motivated to evade).

Both Robert Audi (1982) and Georges Rey (1988) have argued that self-deception does not bring about belief in the usual sense. Audi's (1982, p. 138) view, for example, only requires that the agent be disposed sincerely to avow the relevant proposition. An 'avowal' or 'avowed belief' means a disposition or tendency to endorse a propositional content verbally (either privately or publicly). This view avoids the static puzzle because it takes self-deception to be a conflict between different kinds of attitudes, namely, full-blown beliefs and mere avowals. It tries to explain how a self-deceived subject might sincerely speak about, say, not being cheated by an unfaithful spouse, while retaining a belief in the doxastic alternative. However, it does so by denying that self-deception entails other properties of beliefs proper, such as their deep connections to non-verbal action.

Finally, among the main approaches adopted to explain self-deception, the most widely espoused recently has been the *deflationary* account, marked on the one hand by its rejection of the literalist interpretation (and of the mysterious, homuncular solutions that have been proposed to the puzzles it engenders) and, on the other hand, for its choice of abandoning the ascription of unconscious, inaccessible belief in the doxastic alternative (Mele 2001).[6] According to this view, the mental state and product of self-deception is simply a form of motivated false belief—the subject has only one belief, namely, the belief in the proposition the self-deception is about. That is, a subject in the hold of self-deception is seen as actually believing the false or unwarranted proposition that is the content of his desire (and as not believing the true or warranted proposition for which there is sufficient evidence available). Support for this view comes from the fact that self-deceivers are usually sincere in their assertions; that, upon reflection, they will assert that they believe the relevant proposition; and that they

---

[6] See also Johnston (1988), Barnes (1997), and Van Leeuwen (2007).

JOSÉ EDUARDO PORCHER

will often act (and not only react) on the basis of the content of their self-deceptions, sometimes with dire consequences (while, by contrast, the avowal view is only able to explain verbal behavior).[7] In what follows I will refer to this as the *doxastic conception* of self-deception and will ignore the subtleties of different doxastic explanations to pursue the more fundamental question of whether 'self-deceptive belief' is a tenable notion at all.

2. What's wrong with 'self-deceptive belief'

Doxastic accounts of self-deception have recently been met with criticism from Tamar Gendler (2007). She follows David Velleman (2000) in rejecting explanations of belief that distinguish it from other cognitive attitudes solely on the basis of its role in the motivation of action. In his paper, Velleman produces a number of examples with the aim of showing that other cognitive attitudes, most importantly propositional imagination (i.e. imagining that p), can motivate action in the manner taken by some theorists to be exclusive of belief. His main claim is that, seeing that the motivational role of belief is shared by other attitudes, only the fact that it aims at the truth can successfully distinguish it from the other attitudes (whose aim falls short of truth). Differently from merely surmising or imagining that a proposition is true, believing a proposition consists in bearing the attitude that one does with the aim of thereby accepting a truth. In complete agreement, Gendler claims that belief is inherently 'reality-sensitive' and that it consists in a 'receptive,' as opposed to a 'projective,' attitude. Self-deception, on the other hand, seems to be anything but reality-sensitive. On this basis Gendler (2007, p. 247)

---

[7] Of course, there are also disadvantages to this approach. For instance, this view may fail to account for both the epistemic tension usually thought to be inherent in self-deception and the avoidance behavior characteristic of self-deceivers (Porcher 2012).

argues that if belief is the attitude of accepting a proposition with the aim of thereby accepting something true, then in cases such as that of self-deception our thoughts are occupied and our actions are guided by contents that we do not actually *believe*. This conclusion, she observes, should not lead us to revise and relax our standards for belief but, rather, it should leads to recognize that other cognitive attitudes also play the role most distinctively tied to belief. Here is the structure Gendler's argument against the doxastic account of self-deception:

1.   A mental state is a belief if and only if it aims at the truth. [premise argued for by Velleman (2000)]

2.   Self-deception does not aim at the truth. [premise, self-evident]

3.   Therefore, self-deception is not belief. [from 1 and 2]

Note, however, that Gendler's first premise actually warrants a stronger conclusion. If belief is indeed essentially truth-oriented as she assumes throughout her paper, then not only do we not believe the content of our self-deceptions, but we *cannot* believe it. If her argument succeeds, then, it completely rules out belief as a candidate for the kind of state in which self-deception consists—an impressive achievement considering that it would do away with almost every extant account of the phenomenon.

In reaction to what she perceives as a fatal blow to the doxastic conception, Gendler proposes a novel account of the characteristic cognitive attitude of self-deception.[8] She maintains that self-deception

---

[8] A cognitive attitude is one that treats its content as true; hence beliefs, surmises, assumptions, acceptances in a context, and also imaginings are cognitive attitudes. Their counterparts are conative attitudes, which treat their contents as to be made true; hence desires and wishes are conative attitudes. (I

would be best accounted for by appeal to a form of propositional imaginative *pretense*. Gendler's proposal belongs to the family of accounts that maintain that the self-deceived hold a true, temporarily inaccessible belief, and another false or unwarranted attitude that is not itself belief—thus bypassing the static puzzle. In this, her main predecessors are Audi (1982) and Rey (1988).[9] However, the specifics of her view are unprecedented. Also, she has the further, parallel aim of unmasking what she deems a failure to recognize the philosophical significance of the degree to which attitudes other than belief play a central role in our mental and practical lives (2007, p. 231). Here is Gendler's positive proposal, in a nutshell:

> A person who is self-deceived about ¬*p* pretends (in the sense of *makes-believe*, or *imagines* or fantasizes) that ¬*p* is the case, often while believing that *p* is the case and not believing that ¬*p* is the case. The pretense that ¬*p* largely plays the role normally played by belief in terms of (i) introspective vivacity and (ii) motivation of action in a wide range of circumstances. (2007, pp. 233-4)

---

follow Van Leeuwen 2009 in this formulation, who borrows it from Velleman and Shah 2005). As much as self-deception involves (in most, if not all of its instances) one or more conative attitudes in its formation and maintenance, its end-product isn't itself a conative, but a cognitive attitude. It isn't a representation of how the subject wants the world to be, but a representation of how the world actually is for the self-deceived, and thus it possesses a far-reaching role in both the theoretical and practical reasoning of such subjects.

[9] Audi, Rey and Gendler's self-deceivers resemble what Raymond Smullyan calls *peculiar reasoners*: 'We will call a reasoner *peculiar* if there is some proposition *p* such that he believes *p* and also believes that he doesn't believe *p*. (This strange condition doesn't necessarily involve a logical inconsistency, but it is certainly a psychological peculiarity!)' (1986, p. 344). However, it will be important for the avowal/pretense view that the self-deceiver also have the false second-order belief that she believes that ¬*p*. Smullyan's discussion did not capture this further peculiarity, but perhaps this is just as well, since it is a matter of dispute whether such reasoners do in fact exist (but see Moran 2001).

Consequently, Gendler's self-deceivers do not come to believe the content of their self-deceptions, but engage in a form of mental *simulation*: their motivation to avoid the recognition of some truth or other ($\neg p$) leads them to mentally escape the real world and intermittently inhabit a "*p*-world," an imaginary environment which protects them from the inconvenient or undesired evidence. The usual tension displayed by self-deceivers is explained by allowing what Gendler terms a 'projective' attitude (namely, pretense) to play a role in a context in which rationality would mandate that a 'receptive' attitude (namely, belief) do the work. The avoidance behavior characteristic of self-deceivers, in turn, can be explained by the retention, albeit tacit, of the true or warranted belief. Finally, the threat of contradiction or paradox is relieved by the appeal to something other than belief—a different attitude brought by a different cognitive process.

Nevertheless, whatever Gendler means by 'pretense,' it cannot just be something like the daydreaming kind of fantasy which would seldom, if ever, dispose the subject to sincerely avow its content. In self-deception it is expected that the subject not only evade the harsh truth and a sufficient part of the body of evidence pointing to it— something Gendler's pretenders in their avoidance and mental flight fitfully exemplify—but they also must abide by their self-deceptive attitudes, whatever they are. Hence, as Christoph Michel and Albert Newen (2010, p. 736) point out, to play the explanatory role it is meant to play, Gendler's concept of pretense must be a hybrid that has to be belief-like in explaining the subject's *p*-behavior and *p*-confidence, while being sufficiently imagination-like so as not to conflict with the subject's knowledge that *p* is untrue. For the sake of the argument, however, I will not at this moment object to the promiscuous proliferation of kinds of cognitive states, in order that we may assess the merits of Gendler's hybrid propositional attitude.

Having briefly presented the portions of her original account which are more immediately important to what will follow, it is worth noting that what Gendler advances is very different from, say, allowing a *role* for pretense in the explanation of the *process* of (some forms of) self-deception. Hers is a bold and sweeping claim. According to her, self-deception just *is* pretense, that is, the product of self-deception is pretense (a state that is belief-like but that falls short of constituting full-blown belief). In the next section, I turn to assessing the cogency of Gendler's account by evaluating its presuppositions, which are drawn from Velleman's account of the aim of belief and of the motivational role of belief and other cognitive attitudes.

3. Self-deception as pretense

3.1 Velleman on belief and Gendler's appropriation

In spite of her argument's sweeping conclusion, Gendler (2007, p. 236) explicitly acknowledges the fact that there are numerous ways in which belief can obtain without its normal manifestations, and that it is certainly possible for someone to have false or ill-grounded beliefs. While this may seem obvious, it warrants the following question: if we accept that these statements are true, why can we not characterize self-deception precisely in these terms (which are, in fact, the doxastic theorist's terms)? The answer, embodied in the second premise of Gendler's argument against the doxastic conception, is that in self-deception the aim of the subject's cognitive attitude is not truth—the self-deceived don't bear the attitude that they do with the aim of thereby bearing that attitude toward something true.

However, there are sufficient textual grounds to conclude that Velleman himself would not accept Gendler's second premise. Indeed, he states in a telling footnote that a 'person's cognition of being

Napoleon might . . . remain under the control of truth-directed mechanisms, which were being *diverted* from their goal; and in that case, he would literally have deceived himself, by self-inducing a false belief' (2000, p. 281, my emphasis).[10] This seems to contradict Gendler's claim that self-deception cannot be characterized as belief, since Velleman clearly allows the possibility of self-induced false *belief*. The key element of the quoted passage is Velleman's assertion that, though beliefs may be essentially truth-directed, they may be diverted from their goal. Importantly, this fact must be acknowledged if one is to allow the possibility of *any* kind of misbelief, from simple error to full-blown delusion, with self-deception somewhere along the spectrum. In such cases, someone's cognition may deviate from, and fall short of, its goal for any number of reasons. The moral, I argue, is that it is *extrinsic* to the truth-directedness of belief whether it actually arrives at falsity by whichever means, be it negligence, bias, misinterpretation, etc. In Velleman's own words: 'Faulty or mistaken beliefs are the ones whose regulation has not succeeded in producing the kind of cognitions that it was designed to produce' (2000, p. 278). I suggest that self-deception just is one among many cases where regulation breaks down.

In keeping with the claims I have examined so far, Gendler (2007, p. 245) affirms that pretense is a plausible candidate for the attitude that characterizes self-deception because, if that were a belief, the norms governing it would mandate that it be abandoned on grounds of falsity (or lack of warrant, at least). But note that this would seem to imply that *all* irrational (or otherwise faulty) attitude formation involves some attitude other than belief. In a footnote, however, Gendler seems to point out that she does not aim at such a conclusion:

---

[10] Note, however, that Velleman is actually using an even more dramatic example than self-deception, namely, that of clinical delusion. On the relation between self-deception and delusion, see Bortolotti and Mameli (2012) and the essays in Bayne and Fernández (2009).

'Beliefs formed as a result of cognitive biases . . . may well be defective in certain ways' (2007, p. 254, fn. 46). She follows Ariela Lazar's distinction between attitudes formed out of bias and self-deceptive ones: 'cognitive biases are persistent patterns of biased reasoning. They are exhibited regardless of subject-matter. In contrast, Lazar maintains that self-deception is thematic: the content of the irrational belief is relevant to the explanation of its formation' (1999, p. 267). But Lazar oversimplifies, since a whole family of relevant cognitive biases, so-called 'hot' biases, is characteristically thematic.

Some biases are self-directed, which confers relevance to the content of the subsequent biased belief. Examples are all too familiar. The *self-serving bias* is one such behavioral pattern, in which people attribute their successes to internal or personal factors but attribute their failures to situational factors beyond their control (Miller and Ross 1975). The *Dunning–Kruger effect* is another such pattern, one whereby unskilled people make poor decisions and reach erroneous conclusions, but their incompetence denies them the metacognitive ability to appreciate their mistakes (Kruger and Dunning 1999). The unskilled therefore develop the further bias of *illusory superiority* (Hoorens 1993), rating their abilities as above average, sometimes much higher than they actually are, while the highly skilled underrate their own abilities, suffering from illusory inferiority. Driving, socializing, and solving logic and math problems are common examples of such abilities. This bias provides an explanation to why actual competence may weaken self-confidence—or in Darwin's words, why 'ignorance more frequently begets confidence than does knowledge' (1871/1981, p. 3)—since competent individuals falsely assume that others have equivalent abilities. Kruger and Dunning conclude that 'the miscalibration of the incompetent stems from an error about the self, whereas the miscalibration of the highly competent stems from an error about others' (1999, p. 1127).

Gendler (2007, p. 254, fn. 46) contrasts the biased subject to the self-deceived subject by pointing out that the latter will not readily accept that her attitude toward ¬*p* is unjustified even when this is pointed out to her. The problem is that Gendler compares the self-deceiver to the irrelevant kind of biased reasoner, namely the 'cold' one. A person taken in by a motivated ('hot') form of bias, such as illusory superiority, will arguably exhibit the same, or at least a similar, kind of persistence since the subject-matter is indeed relevant to her attitude.[11] So, ultimately, it seems that Gendler is faced with the following dilemma, being forced to choose between either her hyperbolic conception of truth-directedness, or the possibility of beliefs formed out of motivational bias (and irrational beliefs in general). If one should take the first route, however, it would follow that a person with a tendency to overestimate their positive qualities and abilities and underestimate their negative qualities would not be able to be characterized as really believing the content of her attitude. This further restriction of the domain of false belief seems even more implausible than claiming that the self-deceived don't believe the content of their self-deceptions. Which is of no greater moment than the fact that mistaken subjects actually believe the contents of their misconceptions. Thus, Gendler's approach has the unwelcome consequence that her self-deceivers are never actually mistaken in their attitudes toward the self-deceptive content, which seems incongruent with our everyday experience and evaluation of self-deception in ourselves and in others. What exactly is it that has gone wrong in Gendler's argument so that it entails the impossibility of irrational beliefs? I maintain that what Gendler fails to take into account is the difference between belief (the state) and belief-formation (the process) when talking about truth-directedness—which translates to the difference between the state of

---

[11] On the role that both cold and hot biases play in the formation and maintenance of self-deceptive belief, see Mele (2001, pp. 25-31).

self-deception and the process of becoming self-deceived. While one may hold that it is of the essence of beliefs to be reality-sensitive and to represent our rational commitment to the world, it is patent, as we have seen, that belief-formation can nevertheless find ways to go awry (some of them having something to do with the person's desires). What this means is simply that some of our beliefs are not the product of ideal, perfect rationality—actually, most of them are not (Cherniak 1986). The existence of biased processes of belief-formation does not mean that truth-directedness itself is compromised, nor does it really have anything to do with it. Gendler's argument conflates product and process by referring to the truth-directedness of belief in its first premise and then building on the lack of truth-directedness in the process of self-deception in its second premise.

If what has been said is tenable, then a doxastic account of self-deception is not at all ruled out by accepting Velleman's claim of truth-directedness, but is perfectly consistent with it. I take this to be enough to reinstate belief as a candidate. It does not follow from truth-directedness that beliefs can't be formed in non-truth-directed ways. Hence, it does not follow from truth-directedness that the product of self-deception cannot be, or is not, belief. However, this alone obviously does not speak against Gendler's pretense account. The success of Gendler's positive proposal partially hangs on whether her chosen attitude can do the required work left by giving up belief. If it does, then that would be enough independent support for working with Gendler in building a pretense account.[12] For it to work, however, one needs to show that imaginative pretense can have, as she puts it (following Hume's characterization of belief), the introspective vivacity

---

[12] To justify abandoning talk of belief altogether, however, there must be some advantage in giving it up. As we have seen, that advantage for Gendler is providing an explanation compatible with the view that belief essentially aims at the truth.

and the motivational role of belief. I don't deny that imaginative pretense can indeed be introspectively vivid and will not take issue here with the identity of introspective vivacity between belief and imaginative pretense. I will, however, take issue with her claim that belief and imagination can and do share motivational role, jointly with desires, in the production of action. Gendler does not herself present an argument for the identity of motivational role between belief and pretense, but draws from Velleman, whose argument for this thesis takes up most of his paper. Its importance is crucial in his argument for abandoning motivational role as a distinguishing feature of belief. For Gendler, the thesis is just as important: proving it wrong would require at the very least an emendation of her account. In what follows, an examination of that which Gendler takes for granted will require us to delve deeper in Velleman's argument.

3.2 The motivational roles of belief and imagination

Beliefs, in conjunction with desires, cause and rationalize actions that will make the contents of the desires true, if the contents of the beliefs are true. If I believe there is Marsala wine in my pantry, I will go to it and pick up a bottle, provided I want to drink it or use it in a recipe. Hence the conative contribution to action is pretty straightforward. What about the cognitive side? From experience, I know I have acted on many other, lesser kinds of commitment. I have gone to the store based on the mere surmise that they sell mascarpone cheese. And I have talked out loud while walking down the street based on the mere imagining that I am discussing issues in my relationship with a close friend. So it is safe to say, to begin with, that other cognitive attitudes affect behavior, jointly with conative attitudes, in ways that are similar to the ways belief does. Nevertheless, it is intuitive enough to think that these other cognitive attitudes do not have the same role in producing action.

Let us focus on imaginings. Can an imagining that I have won a hundred million dollars in the lottery and a belief to the same effect be equal in their output? Do I behave in the same way as a consequence of holding each of these cognitive attitudes? It would seem absurd to think so. If I imagine, as in a daydream, that I have won the lottery, I may as a consequence imagine myself buying a luxurious apartment with an ocean view in Ipanema, as well as quitting my job, booking a flight to Tahiti, etc. Depending on the vividness of this mental simulation and the intensity of my desire, I might even go online and check out some real estate websites: not because I am about to actually buy one, but because it might feel good to carry on daydreaming—that is, simulating the experience. On the other hand, if I really believe so, such a degree of conviction has markedly different consequences: not only may I entertain buying that house and quitting my job, I may actually call a real estate agent, make a deposit, announce to my family that I am moving, bid farewell to my colleagues, stop working on this paper, etc. In a follow-up paper to 'On the Aim of Belief', however, David Velleman and Nishi Shah assert that the question of how to differentiate the concept of belief from the concepts of other cognitive attitudes cannot be that belief plays a distinctive motivational role because the motivational role of belief is one that it shares with other cognitive attitudes: 'Assuming that p and supposing that p resemble believing that p in that they dispose the subject to behave as if p were true; and even imagining that p may resemble belief in this respect' (2005, pp. 497-8).

At first sight my little scenario seems to contradict such claims. So far we have produced at least a distinction of degree between believing and imagining with respect to their role in motivating action—that is, we have intuitively created a hierarchy of motivational power and placed belief higher than all other attitudes. Even if one accepts the identity of motivational role thesis, beliefs patently are the

standard background for our actions. This betrays an ambiguity in what 'motivational role' might actually mean. So far it is unclear what may be embedded in the word 'resemblance' in Velleman's and Shah's quote above. After all, what is relevant to our discussion is not whether belief and imagining can have similar effects on behavior at a high level of abstraction. Velleman needs to abide by what Van Leeuwen (2009) has called the *identity of comprehensive motivational role thesis*. This need can be seen once one fleshes out Velleman's argument. To this end, Van Leeuwen provides a very clear formulation of the overall argument of 'On the Aim of Belief':

> 1.   If belief cannot be distinguished from other cognitive attitudes by its role in action output, then it must be distinguished from them by etiology or cognitive input, i.e., regulation and production. [premise]
> 2.   Belief and imagining have the same motivational role, i.e., 'conditional disposition to cause behavior,' a role shared by the other cognitive attitudes as well. [lemma argued for in the paper]
> 3.   Therefore, belief cannot be distinguished from other cognitive attitudes by its role in action output. [from 2]
> 4.   Therefore, belief must be distinguished from other cognitive attitudes by cognitive etiology, i.e., regulation and production. [from 1 and 3]
> 5.   Aiming at truth, i.e., being regulated by mechanisms designed to produce truth in beliefs, is the best candidate among cognitive properties that could distinguish beliefs.

> [intuitive assumption, argued for briefly by
> Velleman in 'Answers to Objections']
> 6.   '... truth-directedness is essential to
> the characterization of belief.' [from 4 and 5]
> (2009, pp. 230-1)

Thus, Velleman argues that, since the motivational roles of belief and imagining (and other attitudes) are shared, we must turn to the notion of truth-directedness to distinguish belief from every other cognitive attitude. Velleman's strategy for demonstrating that 2 is true, as mentioned earlier, consists in pointing out, through a series of examples, that other attitudes besides belief have output in behavior and action. For example, there is make-believe, as when a child pretends that she is an elephant, waving her arm like a trunk, drinking from an imaginary pail of water, etc. There is talking to oneself, as when we walk down the street discussing with an imaginary interlocutor or when we address an imaginary audience as we work on a conference paper. There are psychoanalytic examples, as in a number of cases catalogued and interpreted by Freud, in which a patient behaves and acts motivated by wishful fantasies, as when a jealous child symbolically throws out his baby brother—which was how Freud interpreted Goethe's earliest childhood memory of throwing crockery out a window and watching it smash in the street (1917/1958). And, finally, there is expressive behavior, as in Hume's famous example of the dangling cage in *A Treatise of Human Nature* (I.3.13), where someone who is suspended at a great height trembles with fear and holds on to the bars of the cage, despite acknowledging that she is securely supported.[13]

---

[13] Gendler (2008) describes a similar case taking the behavior expressed by tourists when walking the Grand Canyon 'Skywalk'. She coins the term 'alief'

However, Van Leeuwen (2009, p. 232) observes that, logically, the step from 2 to 3 cannot be made unless it can be shown that belief and imagining have the same motivational role, that is, the same disposition to cause behavior, in a comprehensive sense, in which case 'motivational role' would mean all characteristic effects an attitude of a given kind has on behavior. If Velleman does mean the comprehensive sense, it follows that, other things being equal, imagining that *p* will cause *the same* behaviors as believing that *p*. So the fact that he can logically extract 3 from 2 is of no help: his argument may be logically valid, but it hangs on a premise which is wildly implausible and counterintuitive. On the other hand, it is possible that Velleman means 'motivational role' to be read in a weaker and more abstract way, which Van Leeuwen (2009, p. 232) calls the *vanilla* sense, in which the motivational role of a belief is to cause behavior that will satisfy conations if the belief is true. Velleman's examples indeed serve well the purpose of showing that there are some circumstances in which we act in ways that would make our wants satisfied if the contents of our imaginings were true. But is that enough to demonstrate the truth of 2 (taking the vanilla sense into consideration)? While that is an interesting question, even a positive answer to it would not be enough to save his argument, since it would not be logically valid: from a vanilla interpretation it follows that he has not ruled out the possibility that belief and imagining can be distinguished in terms of action output (something which only the comprehensive interpretation can grant). As Van Leeuwen puts it, if we read him as proposing only a vanilla sense of 'motivational role,' Velleman wins a battle, but ultimately he loses the war.

What exactly are the consequences of this discussion to

and argues that this is a different kind of attitude from the attitudes discussed so far in the literature.

Gendler's explanation of self-deception? We have seen that she does not offer independent argument for the identity thesis. The question now is: does Gendler also need the identity thesis to be read in a comprehensive sense, or can her account succeed with appeal only to the vanilla sense? The answer to this question depends on the specifics of self-deception, that is, on what the output in action of whatever attitude self-deceived subjects hold actually is. It has been shown that identity of comprehensive motivational role is plainly false. But that would not represent a problem for Gendler provided that she abides by the weaker, vanilla sense of 'motivational role' in claiming that pretense can play the role of belief in the motivation of action in a wide range of circumstances.

I think the absence of an all-inclusive clause in her formulation warrants us reading Gendler to mean something other than unrestricted identity of motivational role between belief and imagination. However, her clause 'in a wide range of circumstances' limits the subset of circumstances where identity holds, not the kind of identity. Where most theorists would ascribe full-fledged belief to the self-deceived, Gendler does not. Hence, if she says pretense is what characterizes the mental state of the self-deceived, one can only conclude that, for her, self-deception is the type of case where an attitude besides belief plays the exact same role of, and attains comprehensive motivational role with relation to, belief. Consequently, in order to salvage Gendler's formulation, it must be said that while comprehensive identity across all settings is at the very best implausible, it can nevertheless hold in localized instances (a subset of which is self-deception).

It is of course possible that, notwithstanding the shortcomings of Velleman's argument in proving it, truth-directedness is at least a causal, if not normative feature of beliefs. That is not relevant for my purposes, however, since, as I have previously argued, truth-

directedness does not rule out the possibility that self-deceived subjects actually believe the contents of their self-deceptions. Furthermore, it is perfectly possible that truth-directedness is true *and* that it is possible to distinguish belief from all other attitudes solely on the basis of their motivational role. So far we have seen that Velleman has not properly shown the latter to be false. My present purpose then is to assess whether it is true. As a background to what follows we should briefly review the standard characterization of the motivational view of belief. In Velleman's understanding of it, 'All that's necessary for an attitude to qualify as a belief is that it disposes the subject to behave in ways that would promote the satisfaction of her desires if its content were true. An attitude's tendency to cause behavioral output is thus conceived as sufficient to make it a belief' (2000, p. 255). His many examples are enough to show that this view is overly simplistic and erroneous. Nevertheless, in the face of his main argument's lack of success, it would be prudent not to do away with the motivational view all at once.

The problem we are facing—the challenge of finding a way to properly distinguish belief and imagination—can be traced back at least to Hume's *An Enquiry Concerning Human Understanding*. [14] Hume's solution to the problem, as I see it, would rest on a conception of belief directly attacked by Velleman and Gendler. On Hume's view, belief is 'that act of mind which renders realities more present to us than fictions, causes them to weigh more in the thought ... gives them

---

[14] See, for example: 'Wherein, therefore, consists the difference between such a fiction and belief? It lies not merely in any peculiar idea, which is annexed to such a conception as commands our assent, and which is wanting to every known fiction. For as the mind has authority over all its ideas, it could voluntarily annex this particular idea to any fiction, and consequently be able to believe whatever it pleases; contrary to what we find by daily experience' (V.II).

superior influence on the passions and imagination . . . and renders them the governing principles of all our actions' (V.II). Gendler invokes this passage at the beginning of her paper and aims at making an example out of Hume, so to speak, by showing that successfully demonstrating self-deception to be characterized by an imaginative attitude proves he was ultimately wrong. 'Usurpers,' says Gendler, 'do not always deserve the title of the one whom they usurp' (2007, p. 247). Furthermore, she thereby aims to denounce a widespread ignorance of the role played by a range of other attitudes in the cognitive economy of humans. While I strongly agree, being averse to any oversimplification in our understanding of human cognition, I think there are a couple of things that can be said in favor of a generally Humean view of belief and imagination—the view that belief is, among other things, 'the governing principle of all our actions': a feature that distinguishes it from every other cognitive attitude, a role no usurper can ever play, but only mimic to a degree.

So far we have noticed that there is something missing in the standard characterization of the motivational view, and we know such a view cannot be true. On the other hand, as Lucy O'Brien rightly notes, there is something missing also in Velleman's view that imaginings can play the motivational role of beliefs. 'The attitude of imagining that p, by itself and relative to a fixed background of desires, does not dispose the subject to behave in ways that would promote the satisfaction of his desires if its content were true' (2005, p. 58). That is to say, comprehensive identity of motivational role is false. Also, says O'Brien, 'it seems to be a quite general point that any 'regarding as true' [i.e. cognitive] states which are not beliefs, will require [a] kind of connection to the subject's beliefs about his actual world if they are to result in action' (ibid.). This observation sets the tone for the rest of the discussion in this section, namely, that what must be appended to the standard account is an account of the particular relation in which

beliefs stand to other cognitive attitudes, a relation that (sometimes) confers upon the latter the capacity of producing output in behavior and action. Drawing on work by Bratman (1992), Van Leeuwen (2009) takes strides toward a unifying account that exhibits what is needed for an adequate explanation of what the motivational role of belief is.

3.3 Context and practical ground

The main problem for the motivational view is that every cognitive attitude is apt to play some role in the motivation of action. We do act on the basis of surmises, as when we want to eat something and get up to look for it in the fridge, assuming (but not quite sure) that there is something there. It might be argued that this is nothing but a degree of belief this side of certainty. Even so, and more relevant to the present discussion, we also act on the basis of imaginings, as when we play a game of make-believe and swiftly dodge an imaginary sword, despite the fact that our friend is just putting his hands together *as if* he were carrying a sword.

To provide an answer to Hume's problem, we may begin with Bratman's examination of the difference between belief and what he calls 'acceptance in a context'. According to Bratman, an agent's beliefs provide the default cognitive background for further deliberation and planning. This cognitive background is context-independent, but practical reasoning admits adjustments in what one takes for granted in a specific practical context. 'To be accepted in a context is to be taken as given in the adjusted cognitive background for that context' (1992, pp. 10-11). Van Leeuwen (2007, p. 434) suggests extending Bratman's idea to include all other non-belief cognitive attitudes, so that non-belief cognitive attitudes would require specific contexts in order to function as the background of deliberation for the constitution of action. This implies that whenever an imagining prompts action, it does

so only *in the context* of a game of make-believe (or whatever kind of imaginative play is at work). Likewise, whenever a surmise prompts action, it does so only in the context of an investigation. The motivational power of beliefs, however, is context-independent. For this reason, beliefs are the default cognitive background for the constitution of action.

An extension of Bratman's idea, Van Leeuwen's *practical ground thesis* states that belief is the practical ground of all other non-belief cognitive attitudes in circumstances wherein the latter prompt action. This is to say that, while one may have the impression that non-belief cognitive attitudes prompt action on their own (as Velleman and Gendler), they do so only in virtue of being grounded on belief. Beliefs are, to use Ramsey's (1931, p. 238) metaphor, maps by which we steer. Furthermore, Van Leeuwen (2009, p. 239) adds that beliefs determine if one is in the right *setting* for acting on the basis of another attitude or not. Thus, he observes that beliefs are not only the maps by which we steer, but also the maps by which other maps are chosen and appraised.

Van Leeuwen presents the practical ground relation as the conjunction of three types of relation that can hold between classes of cognitive attitudes:

> 1.   Attitudes of type X are available for motivating actions across all practical settings, while attitudes of type Y depend on the agent's being in a certain practical setting to be effective in influencing action.
> 2.   Attitudes of type X represent the practical setting one is in such that one acts on attitudes of type Y on account of being in that setting.

3.   Attitudes of type X are the cognitive input into choosing to act with attitudes of type Y as input into practical reasoning, when one does so choose. (2009, p. 226)

Therefore, X is the practical ground of Y just in case all three relations hold, and this is precisely the case for the ordered pair <Belief, Acceptance in a Context> and, by extension (Van Leeuwen claims), for the ordered pair <Belief, Imagining>—but never for <Acceptance in a Context, Belief>, <Imagining, Belief> etc.

Van Leeuwen does us the favor of fleshing out his argument for the practical ground thesis in explicit and logically clear form. It begins with the already argued-for premise that the identity of comprehensive motivational role is false, from which it follows that there are practical settings in which the behavioral consequences of imaginings differ from that of beliefs. [15] Moreover, Van Leeuwen argues that these differences in output are non-accidental, since practical setting typically influences behavior in conjunction with cognitions. It follows that they are caused, at least in part, by differences in the very practical settings in which they occur, which implies that either the psychomechanical efficacy of imagining is practical setting-dependent, or the psychomechanical efficacy of belief is, or both. [16] Given that the psychomechanical efficacy of belief is not context-dependent (as Bratman argues), the psychomechanical efficacy of imagining is. From this pair of statements Van

---

[15] Here and for the remainder of the argument, 'imaginings' is meant as shorthand for 'imaginings and other non-belief cognitive attitudes'.
[16] 'Psychomecanical efficacy' is Van Leeuwen's term for the property of being effective in influencing action.

Leeuwen extracts the first of three lemmas in his argument:

> **Lemma 1.** Beliefs are effective in practical reasoning and motivating actions in a practical setting-independent way, while imagining depends on practical setting to be effective in influencing action.

Short of a 'magical connection' between being in a practical setting and having that practical setting activate a specifically adjusted cognitive background for it, it must be assumed that the fact of practical setting-dependence mandates that the agent *represent* the practical setting she is in (Van Leeuwen 2009, p. 237, fn. 19). Given the practical setting-dependence of imagining, it follows that an agent who acts on the basis of imaginings must have a representation of the practical setting she is in. Now, whatever they are, representations of practical setting are certainly cognitive, as opposed to conative, attitudes. However, given that non-belief attitudes are themselves practical setting-dependent, representations of practical setting must be practical setting-*independent* (since otherwise we would have an infinite regress of representations of practical setting). This, in conjunction with the first lemma, yields a second lemma in Van Leeuwen's argument:

> **Lemma 2.** There are beliefs that represent the practical setting an agent is in, on which the psychomechanical efficacy of imagining is dependent.

Now, one of the important conclusions Van Leeuwen (2009, pp. 227–9) derives from a story about him and his childhood friend Chris—which illustrates the third type of relation needed for the practical ground relation to hold—is that acting with imagining as the adjusted cognitive background is a *choice* (whereas being taken in by imaginative play is often involuntary). In short: while playing a game of make-believe in the mud, his friend gets stuck. Initially, the protagonist believes that this is part of the game and carries on. But once his friend informs him that he is really stuck, he turns to acting on other beliefs,

for instance, the belief that by helping him he would become unstuck (as a consequence of which he goes over to get Chris unstuck). Believing, then, that Chris was successfully unstuck, and that the game of make-believe could be resumed, the protagonist makes the choice (against the background of these beliefs) to resume it. Given that attitudes of the kind that represent the practical setting an agent is in are cognitive inputs into choices the agent makes in those settings, a third and final lemma is derived, whence the conclusion of the argument is finally extracted:

> *Lemma 3*. Beliefs are cognitive inputs into choosing to act with imaginings as the adjusted cognitive background, when one does so choose.
> *Practical ground thesis*. Belief is the practical ground of imagining. [from lemma 1, lemma 2, lemma 3, and the definition of practical ground ...] **QED**

This concludes the gloss on Van Leeuwen's argument. I now turn to filling in the gaps in Gendler's account by looking at how Van Leeuwen's results affect it.

3.4 The practical ground of self-deception

Can imaginative pretense, or fantasy, or make-believe sometimes play the role that belief does in the motivation of action? So far we have seen that pretense and other non-belief cognitive attitudes can and do play a role in motivating action in a wide range of cases. On the other hand, we have learned from Bratman and Van Leeuwen that this doesn't warrant ascribing them the same role. If the practical ground thesis is true—that is, if belief is the practical ground of imagining—then every time we perform an action on the basis of an imagining, we have beliefs that represent the practical setting we are in, on which the psychomechanical efficacy of said imaginings depends.

This presents a problem for Gendler's account since self-deception requires a certain (variable) level of ignorance of one's own situation. Provided that Gendler does claim that imaginings can be psychomechanically effective (and are so in cases of self-deception), acceptance of Van Leeuwen's practical ground thesis would imply that, in self-deception, the agent has beliefs that represent this peculiar type of practical setting. While this is compatible with a general theory of make-believe, it means that self-deceived agents, to act as they do on their supposed pretenses, are required to believe that they are pretending, or fantasizing, or making-believe. However, it is plainly impossible to be motivated to act on our self-deceptions if not only do we not believe their content, but we simultaneously believe that we are only pretending that such content is true (or, to use Gendler's terms, that we are in a world where what we want to be true is true).

Another difficulty arises by the application of Van Leeuwen's framework to Gendler's account, since acting with imagining as the adjusted cognitive background is supposedly a choice. I take it that this does not imply that we cannot break into tears while watching a film—for example, imagining that James Stewart's character in *It's a Wonderful Life* really is delivering a heart-wrenching speech—or mumble something as we walk down the street—imagining that we are finally letting our friend know what we think of his drinking habit. These are things we do while daydreaming or so and we do them involuntarily. Someone who voluntarily speaks his mind out loud while walking down the street is not in the grasp of imaginative pretense but something altogether different. What Van Leeuwen claims is that voluntary action in the adjusted cognitive background of imagining, such as when we make-believe we are conductors directing Beethoven's Ninth, is a choice. The question then is, does Gendler have in mind the mumbling-down-the-street, involuntary type of action when she describes the workings of self-deception? Or does she mean the

waving-hands-as-if-conducting, voluntary type of action? While one hardly could mean the first sense, it is worth delving briefly into everyday cases of self-deception to see more clearly why one could not.

First, consider a businessman who overestimates his own skills, underestimates (and often just ignores) the failure of his past and current enterprises, and who is very much in debt. He often talks about opening up new businesses, much to the dismay of his family, and has a knack for finding improbable ventures and locations, such as selling car batteries in a beach town whose population drops from 30,000 in the summer to 3,000 throughout the rest of the year. Nevertheless, each time he seems firm in his conviction that his business ideas will prove successful and lucrative (and each and every time he is proven wrong). He is not just stubborn, but adamant, and won't listen to reason, won't extract from the evidence the same conclusion anyone else would, and won't heed the advice of friends and family. He is self-deceived, and his most recent course of action while in the tight grasp of this state of mind was to request a considerable loan. Cases like these are commonplace. It is not in the best interest of people like the businessman to perpetuate and plunge even further into his already outstanding debt. But he deliberately does exactly that.

Second, consider a case where the misfortune lies not so much in acting, but in deliberately failing to act. A single mother who welcomes a man into her home and, despite the heaping evidence that her daughter's new stepfather is crossing the line between innocent and lustful affection, refuses to acknowledge that he might be sexually interested in her daughter. This is also a situation which is not uncommon and that can have appalling consequences. Let me elaborate. The mother is not subject to full-blown delusion. She does notice that the way her boyfriend treats her daughter is increasingly

aggressive. She also notices that her daughter's face shows great distress when she is asked about the subject, and that her behavior has rapidly changed from that of a docile to that of an injured and indignant child. Again, anyone else, given the same amount of evidence she has, would be quick to conclude that if something grim has not already happened, it is about to. But the mother does nothing. She refrains from asking any more questions, since just thinking about the subject upsets her a great deal. However, she is not at all devoid of motherly love: she is just blindly in love with the man, who responds with violent indignation to the mere hint of her preoccupation. As with the businessman, her actions are the product of self-deception, but differently from my previous scenario, the mother sins by omission rather than by actually doing something.

Can deliberate courses of action like these be taken on the basis of mere imaginative pretense? Van Leeuwen (2007) asks a similar question, but he aims it at avowal, rather than pretense. As I have said, Gendler's account bears some resemblance to those that appeal to avowal, since these take the doxastic alternative to be the object of a tacit or unconscious belief, while claiming that the product of self-deception is not itself belief, but a weaker attitude. In that article, Van Leeuwen criticized the avowal view, arguing that it is absurd to propose that action patterns as serious as those just now portrayed could happen as a consequence of holding something short of full-fledged belief. The way Audi and Rey define avowal, however, justifies Van Leeuwen's criticism in such a way that we cannot simply apply to Gendler's view: for them, an avowal is different from a proper belief because it lacks belief's connection to action, whereas Gendler claims imaginative pretense *does* have such a connection. So the avowal view simply cannot explain the import of an action such as taking out a loan at great personal risk instead of cutting losses.

Whether or not imaginative pretense can account for the kinds of action performed on the basis of self-deception, we gather from examples such as those I have just now discussed that, in forging a model of self-deception, one must account for the actions that make self-deception a serious and potentially hazardous issue for the people involved. We have seen that one can easily choose to act, as in a game of make-believe, with imagining as the adjusted cognitive background. The relevant representation, which might be brought to consciousness with the form 'I believe that I am pretending that I am William Wallace' (or whatever) has consequences such as deliberately moving one's arms about as if one is carrying a heavy sword, yelling 'FREEDOM!,' etc. On the other hand, if Van Leeuwen is right, then it follows that one simply *cannot* really act on self-deceptive, imaginative pretense. This is so because *choosing* to act on self-deceptive pretense would constitute nothing short of a self-defeating project, in the fashion of the old dynamic puzzle (Mele 2001), since imaginative pretense can motivate action only insofar as it is backed by a meta-representation—namely, a belief about the setting. But the only way a self-deceived agent could act on her self-deceptions would be by being completely oblivious of the practical setting she is in: 'I believe that I am pretending that my girlfriend is faithful' would never have the kind of consequences it would have in cases where the subject is truly self-deceived, such as behaving as if nothing is wrong most of the time, asking her to move in (especially since the subject's central belief, according to Gendler, would be of the form 'I believe my girlfriend is unfaithful'). The only way one can come to act on the basis of self-deception is to falsely or unwarrantedly *believe* that its content is true, however unjustified that might be.

4 Conclusion

By analyzing the presuppositions of Gendler's account—embodied in her interpretation of the truth-directedness of belief and on her reliance on the identity of motivational role between belief and other cognitive attitudes—I have argued for two claims which put considerable pressure on the idea of self-deception as imaginative pretense. First, I have argued that Gendler's use of belief's alleged essential truth-directedness stems from confusing the rationality of belief formation (a process) and the rationality of belief *per se* (its product). While belief as a cognitive attitude may be said to essentially aim at the truth, the processes by which beliefs are formed are obviously fallible, being subject to a great variety of missteps. I argued that attention to this fact suffices to make clear that what happens in cases of biased cognition (be it hot, cold, or both), including self-deception, is not a violation of the truth-directedness constraint that would warrant the abandonment of belief talk altogether in this context. Truth-directedness can be held to be a feature of belief even in the context of biased cognition, where belief is formed and maintained in ways that divert it from its ideal, rational aim. In sum: Gendler's argument against the doxastic conception of self-deception fails.

Second, as Gendler's account depends on the success of Velleman's argument to the effect that belief and imagination have the same motivational role—a very counterintuitive hypothesis, as I have tried to show in a few scenarios—I have drawn on Van Leeuwen's analysis of Velleman's argument. As Van Leeuwen skillfully shows, Velleman conflates a weaker and a stronger sense of 'motivational role' and, moreover, the success of Velleman's argument depends on the stronger sense, which would commit him to the wildly implausible view that imagining that *p* has *the same* motivational power of believing that *p*. On the other hand, Van Leeuwen's positive account of the relation

between belief and other cognitive attitudes culminates in the practical ground thesis: belief is the practical ground of all other non-belief cognitive attitudes in circumstances wherein the latter prompt action. Drawing on this, I have applied Van Leeuwen's framework to Gendler's account and showed that it engenders a variety of the dynamic puzzle of self-deception: in order to voluntarily act in the context of imaginative pretense, the self-deceiver would have to know that she is in the context of imaginative pretense and choose to act on the basis of such pretense—and this, in turn, would make for a self-refuting project in the case of self-deception. In sum: Gendler's argument for self-deception as pretense relies on a problematic argument for the identity of motivational role and, if Van Leeuwen's practical ground thesis holds, her account cannot make sense of voluntary actions that are a commonplace result of self-deception.

Note, however, that although I have argued against the view that self-deception is imaginative pretense, it is still possible (and plausible) that pretense figures in the process of forming and maintaining a self-deceptive belief and, thus, it may be given a role in an explanatory account of self-deception. Mike W. Martin acknowledged 'self-pretense' among the many patterns of evasion seen in self-deception (1979; 1986). It is in Martin's way that I think we should incorporate Gendler's (2007, p. 240) core insight, namely, that just as you can deceive another by performatively pretending that $\neg p$ rather than p, so too one can deceive oneself by imaginatively pretending that $\neg p$ rather than p. What I have objected to here is taking imaginative pretense to be the *product* of self-deception instead of one of the ways through which people can deceive themselves. And I have tried to do that by showing that Gendler's appropriation of Velleman's theory of truth-directedness is misguided and that, in addition, Velleman's theory of the motivational role of non-belief cognitive attitudes (on which Gendler's explanation depends) doesn't stand up

under close scrutiny.

REFERENCES

AUDI, R. Self-Deception, Action, and Will. *Erkenntnis* 18: 133–58, 1982.

BAYNE, T. and FERNÁNDEZ, J. *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*. Psychology Press, 2009.

BARNES, A. Seeing through Self-Deception. Cambridge University Press. In: Bayne, T. and Fernández, J., (eds.) (2009). *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*. Psychology Press.

BORGE, S. The myth of self-deception. *Southern Journal of Philosophy* 41: 1–28, 2003.

BORTOLOTTI, L. and Mameli, M. Self-Deception, Delusion and the Boundaries of Folk Psychology. *Humana.Mente*, 20: 203–221, 2012.

BRATMAN, M.E. Practical Reasoning and Acceptance in a Context. *Mind* 101 (401): 1-16, 1992.

CANFIELD, J.V. and Gustafson, D.F. Self-Deception. *Analysis* 23: 32–36, 1962.

CARSON, T. L. Lying, Deception, and Related Concepts. In: C.W. Martin, (eds.) (2009), *The Philosophy of Deception*. Oxford University Press.

CHERNIAK, C. *Minimal Rationality*. MIT Press, 1986.

CURRIE, G. and RAVENSCROFT, I. *Recreative Minds: Imagination in Philosophy and Psychology.* Oxford University Press, 2002.

DARWIN, C. *The Descent of Man, and Selection in Relation to Sex.* Princeton University Press, 1871/1981.

DAVIDSON, D. Deception and Division. In: J. Elster, (eds.) (1985), *The Multiple Self.* Cambridge University Press.

DAVIES, M. Delusion and Motivationally Biased Belief: Self-Deception in the Two-Factor Framework. In: T. Bayne and J. Fernández, (eds.) (2009) *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation.* Psychology Press.

DEMOS, R. Lying to Oneself. *Journal of Philosophy*, 57: 588–95, 1960.

ELSTER J. *The Multiple Self.* Cambridge University Press, 1985.

EVANS, J. St. B. T. and FRANKISH, K., (eds.) (2009). *In Two Minds: Dual Processes and Beyond.* Oxford University Press.

FREUD, S. A Childhood Recollection from Dichtung und Wahrheit. In: J. Strachey et al., (eds.) (1917/1958), *The Standard Edition of the Complete Psychological Works of Sigmund Freud.* Hogarth Press.

GENDLER, T.S. Self-deception as pretense. *Philosophical Perspectives* 21 (Philosophy of Mind), 231–258, 2007.

GENDLER, T.S. Alief and Belief. *Journal of Philosophy* 105, 634–663, 2008.

GERGEN, K. The Ethnopsychology of Self-Deception. In: M. Martin, (eds.) (1985), *Self-Deception and Self-Understanding.* University of Kansas Press.

HAIGHT, M. *A Study of Self-Deception.* Harvester Press, 1980.

HOORENS, V. Self-enhancement and Superiority Biases in Social Comparison. *European Review of Social Psychology*, 4(1): 113–139,

1993.

JOHNSTON, M. Self-Deception and the Nature of Mind. In: B. McLaughlin and A. Rorty, (eds.) (1988), *Perspectives on Self-Deception*. University of California Press.

KAHNEMAN, D. *Thinking, Fast and Slow*. Farrar, Straus and Giroux, 2011.

KANT, I. *The Metaphysics of Morals*, trans. M.J. Gregor. Cambridge University Press, 1797/1996.

KIPP, D. On Self-Deception. *Philosophical Quarterly*, 30: 305–17, 1980.

KRUGER, J. and DUNNING, D. Unskilled and Unaware of It: How Difficulties in Recognizing One's Own Incompetence Lead to Inflated Self-Assessments. *Journal of Personality and Social Psychology*, 77(6): 1121–34, 1999.

LAZAR, A. Deceiving Oneself or Self-Deceived? On the Formation of Beliefs 'Under the Influence.' *Mind* 108: 265–90, 1999.

MARTIN C. W. *The Philosophy of Deception*. Oxford University Press, 2009.

MARTIN, M. W. Self-Deception, Self-Pretence, and Emotional Detachment. *Mind* 88: 441-446, 1979.

_____. *Self-Deception and Self-Understanding*. University of Kansas Press, 1985.

_____. *Self-deception and Morality*. Kansas University Press, 1986.

MCKAY, R., LANGDON, R. and COLTHEART, M. "Sleights of Mind": Delusion and Self-Deception. In: T. Bayne and J. Fernández, (eds.) (2009) *Delusion and Self-Deception: Affective and Motivational Influences on Belief Formation*. Psychology Press.

MCLAUGHLIN B. and RORTY A. *Perspectives on Self-Deception*. University of California Press, 1988.

MELE, A.R. *Self-Deception Unmasked*. Princeton University Press, 2001.

MICHEL, C. and NEWEN, A. Self-Deception as Pseudo-Rational Regulation of Belief. *Consciousness and Cognition* 19 (3): 731-744, 2010.

MILLER, D. T. and ROSS, M. Self-serving biases in the attribution of causality: Fact or fiction? *Psychological Bulletin*, 82, 213-225, 1975.

MORAN, R. *Authority and Estrangement*. Princeton University Press, 2001.

NEU, J. Life-Lies and Pipe Dreams: Self-Deception in Ibsen's The Wild Duck and O'Neill's The Iceman Cometh. In *A Tear Is an Intellectual Thing*. Oxford University Press, 2000.

O'BRIEN, L. Imagination and the motivational view of belief. *Analysis* 65 (285): 55–62, 2005.

PEARS, D. *Motivated Irrationality*. Oxford University Press, 1984.

PORCHER, J. E. Against the Deflationary Account of Self-Deception. *Humana.Mente*, 20: 67–84, 2012.

RAMSEY, F. P. *The Foundations of Mathematics and Other Logical Essays*. Routledge, 1931.

REY, G. Toward a computational account of akrasia and self-deception. In: B. McLaughlin and A. Rorty, (eds.) (1988), *Perspectives on Self-Deception*. University of California Press.

SARTRE, J. P. *Being and Nothingness*, trans. H. Barnes. Pocket Books, 1949/1957.

SMULLYAN, R. M. Logicians who reason about themselves. In *Proceedings of the 1986 Conference on Theoretical Aspects of Reasoning*

about Knowledge. Morgan Kaufmann, 1986.

STRACHEY J. et al. *The Standard Edition of the Complete Psychological Works of Sigmund Freud*. Hogarth Press, 1917/1958.

VAN LEEUWEN, D. S. N. The Product of Self-Deception. *Erkenntnis* 67(3), 419–437, 2007.

_____. The Motivational Role of Belief. *Philosophical Papers* 38(2), 219–246, 2009.

VELLEMAN, J. D. On the Aim of Belief. In *The Possibility of Practical Reason*. Oxford University Press, 2009.

_____. and SHAH, N. Doxastic deliberation. *Philosophical Review* 114: 497–534, 2005.