

What Does the Self-Deceiver Want?*

Patrizia Pedrini[†]
patpedrini@gmail.com

ABSTRACT

According to a recent theory of the motivational content of self-deception (Funkhouser, 2005), the self-deceiver wants to be in a state of mind of belief that p , upon which her want that p be true would be merely contingent. While I agree with Funkhouser that the self-deceiver is considerably moved by an interest in believing that p , which makes it possible for her to relate to reality in a highly prejudiced way, I will argue that it is unlikely that the self-deceiver's primary want to believe, or interest in believing that p occurs as the result of a merely contingent interest in p being true. I will finally assess various consequences of the view I favor, regarding the self-deceiver's avoidance behaviour, "twisted" self-deception, and whether we should provide a unifying account of "straight" and "twisted" self-deception.

* I am most grateful to Eric Funkhouser and Dana K. Nelkin for crucial discussions on an earlier version of this paper, and for their pertinent and generous comments; to Akeel Bilgrami and Achille Varzi for many exceptionally stimulating conversations, during my two years at Columbia University in New York City, about my interest in self-deception, among other topics, and the route I was even then showing an inclination to take; to Alessandro Pagnini for having welcomed and encouraged over the years my interest in self-deception, and to Tito Magri for agreeing to publish my second book in Italian on this topic; to the audience in Milan, at the 2011 European Conference of Analytic Philosophy, where a shorter version of this paper was accepted and presented; to the incredibly good audience at the University of Rijeka Department of Philosophy, where I served two semesters as fixed-term professor and, in November 2011, gave a senior lecture that allowed me to discuss in depth one of the last versions of the paper. I also wish to thank the Italian Mensa Society formally for offering me one-year funding to work on what will hopefully be my first book in English on the topic of self-deceptive motivation.

[†] Senior Research Fellow in Philosophy funded by the Mensa Society; Fixed-Term Professor, Dept of Philosophy, University of Rijeka; Assistant Fellow, College of Letters and Philosophy, University of Modena & Reggio-Emilia, and Dept. of Philosophy, University of Florence.

What does the self-deceiver want? Does she want to reach a state of mind, that is, the belief that p , which she likes, or wants, to believe, or does she also want reality to be exactly as she wants it to be, that is, p to be *true*? In other words, what is the operative desire that leads her to self-deception? Does she have a *self-focused desire* to be in the state of mind of belief that p , regardless of whether p is true or false, or does she have a *world-focused desire* that p be true? According to a recent theory of the motivational content of self-deception (Funkhouser, 2005), the self-deceiver wants to be in a state of mind of belief that p , which she finds pleasant or anyway “interesting”, or “important” for her to be in, while not necessarily also being focused on the task of dispassionately ascertaining how things actually stand in the world. On this account, the self-deceiver is thought to have a self-focused desire (Funkhouser, 2005, p. 296) to be in a certain state of mind of belief that p , or a “desire to believe” (Nelkin, 2002) that p , and this would be the leading motivation for self-deception. She can contingently have a world-focused desire (Nelkin, 2002, p. 296) that the world be such that p be true, but such desire is not intrinsic to the self-focused motivation for self-deception and in fact is sometimes lacking. The contingency of her world-focused desires that p be true are demonstrated by her “avoidance behavior”, according to Funkhouser: typically, the self-deceiver actively avoids evidence suggesting that p is false (2005, pp. 297–298). Were she to be dispassionately interested in the truth-value of p and in believing what is true, she would not avoid such evidence. Furthermore, according to Funkhouser, the self-focused desire account of the self-deceptive motivation has the advantage, as we will see, of unifying two kinds of self-deception that the traditional world-focused desire accounts cannot explain: “straight” and “twisted self-deception”. It also has the consequence of helping us to single out a phenomenon that Funkhouser dubs “apathetic” or “indifferent” self-deception (2005, p. 298), which is described as a kind of self-deception where there is *no* world-focused desire of a contingent kind either, but just a self-focused “desire to believe” that p .

In what follows, I will agree with Funkhouser that the self-deceiver is very much attracted by the qualitative aspects of believing that p , and that she may be considerably moved by an “interest” in believing that p , which makes it possible for her to relate to reality in a highly prejudiced way, but, contrary to Funkhouser, I will argue that it does not seem likely that the self-deceiver’s want to believe, or interest in believing that p occurs as the result of a merely

contingent interest in p being true. On closer inspection, it transpires that the self-deceiver finds it pleasant, or interesting, or important, as the case may be (as we will see, depending on the kind of self-deception she embarks on), to believe that p and is attracted to it *exactly because* she wants p to be true, and so she does not embrace the self-deceptive belief that p in a way which could in principle be considered as independent of the interest in being p true, which I take to be *fundamental* and not merely contingent. Considering the self-deceiver's interest in p being true as merely contingent significantly underdescribes the psychological complexity of the motivation that triggers self-deception; furthermore, it leads to counterintuitive conclusions on the very nature of self-deception, which could be easily confused with scenarios of false beliefs "artificially self-induced", where, however, we lose the grip on important specificities of the phenomenon of self-deception and its motivation that are *just* linked to the self-deceiver's world-focused desire, or want, more generally, that p be true.

At the same time, there is an alternative account of the self-deceiver's avoidance behaviour that does not force us to conclude that she is merely contingently focused on wanting reality to be such as p , but rather mainly focused of wanting to believe that p . Typically, the self-deceiver actively looks for evidence suggestive of the truth of p , while avoiding the evidence suggesting that p may be false. Also, she generally gives some treatment of the sources of evidence that p may be false (typically, she does so in a motivationally biased way), as opposed to just avoiding the evidence against p altogether. She typically explains to herself why such sources of evidence are not worth attending, and this epistemic work is the symptom that she is interested in the way the world is and in the truth-value of what she believes. All avoidance behavior shows, as we will see, is that the interest in p being true is strongly biased by the interest in believing that p , but it does not also show that the interest in p being true is not the fundamental engine of the very desire, or want, to believe that p . For these reasons, as I will explain, one could even doubt that "apathetic" or "indifferent" self-deception exists at all.

Finally, I will critically assess the prospects of the unified account of "straight" and "twisted" self-deception reached by defenders of the self-focused desire account and, in this connection, I will say a few words about why some people think that cases of twisted self-deception offer the strongest support to the "desire-to-believe" account of the self-deceptive motivation. I think the question of whether twisted self-deception really offers this support

is, at best, still unsettled, and I will try to suggest considerations that might be useful for a hopefully more decisive future defence of a world-focused approach to the motivation for self-deception.

1. Can the self-deceiver's world-focused desire, or interest, that p be true be *merely contingent* upon her self-focused desire to believe, or interest in believing, that p ?

It is a widespread intuition that when we believe some proposition p , we also, intrinsically, believe that p is true. Believing that p , on this view, is just taking p as true; conversely, if we take p to be true, we, by the same token, believe that p . That is, believing is said to be truth-oriented and the aim of belief seems to many to be that of representing the world as the believer takes it to be. Thus, one who submits to such a view on what believing is, might easily conclude that if one is motivated to believe that p , one must be also intrinsically motivated to take p as true, and that the motivation to believe that p is intelligible in the light of the motivation to take p as true. As far as I can see, however, there are cases where the two kinds of motivation can part company. The most obvious case is the case of the motivational set of someone who wants to acquire an “artificially self-induced” false belief or other mental states. Under the heading of “artificially self-induced” false belief or other mental states, more generally, I do not necessarily refer to Matrix-scenarios such as the voluntary implantation of a belief or a mental state *via* the use of futuristic computerised machines. I also include more ordinary cases of voluntary acquisition of mental states *via* everyday, do-it-yourself “techniques”. These techniques may generally include drugs, alcohol, and other addictive substances, for instance. We are all familiar with real or fictional subjects who strongly wish to cut themselves off from reality, at least for a while: they may just want temporarily to forget how reality is and what they believe it is like; or they may want to acquire a joyful mood to replace their beliefs and/or other unpleasant states of mind; or they may just want to experience what it is like to become convinced, by means of the stimulating effects of a substance on their cognition and memory, that a certain belief that p is true. It may be an issue how exactly those psychological mechanisms actually work, and what exactly the causal chains, initiated by the substance that lead us to acquire a belief that we ultimately want, must be. Also, it is far from clear that using substances can always lead us to believe exactly what we would like to believe. It strikes many people that often it does not, but

we can agree that sometimes it can, and I suggest, for the sake of argument, focusing on the very case in which this outcome is successfully achieved. If this artificial achievement is possible at all, however causally complex and descriptively unclear it may be, we can say that people who successfully reach their doxastic goals via any of these causal chains achieve *artificially* what they could not achieve *epistemically*: since they cannot become convinced, before they enter the self-induced artificial causal chain, that p is true, given the evidence at their disposal, they try to reach that conviction by artificially modifying their perception, or cognition, of the available evidence, forgetting it altogether, altering their reasoning, and so on. Be that as it may, the crucial point of the illustration is that the motivational set of these subjects seems to demonstrate that they have a self-focused desire to acquire, or interest in acquiring, a certain state of mind, including beliefs, while lacking the world-focused desire, or interest, that those states of mind, particularly the beliefs they successfully reach, be representative of the world. They want to enter for a while an inauthentic representation of reality where it seems to them that reality is different from what it actually is, and if the outcome of the artificial modification of their cognition successfully matches their interests, they end up believing that reality is exactly as they want it to be. Note that some may presumably do so because they would like reality to be different, but such world-focused interest does not seem to be intrinsic to the motivation to acquire the artificially self-induced belief. For not only does it seem reasonable to say that substance users, who are motivated to acquire the belief that p because they would like p to be true in some possible, fictional world, are not also necessarily motivated to *establish* that p is true in the real world, as they just seem to want to *believe* that p simply for the sake of the pleasure, where appropriate, or, more generally, satisfaction (which, as we will see, may even be accompanied by discomfort) they get in so believing, regardless of the truth-value in the real world of the proposition they come to believe; also, we could easily imagine subjects who experiment with a substance just because they want to acquire a certain belief that p which it is important for them to acquire, without having any contingent interest in p being true, or any sense that p being true really matters to them. That is, one may want to feel what it is like to believe that p for a while, because it is important to believe so, without having any particular preference regarding p being true at all in the real world. As an example, assuming that there is a technique for coming to believe that one is a brilliant mathematician, it may be pleasant, or important, or interesting to

believe for a while that one is a brilliant mathematician, and thus one could be motivated by the qualitative aspects of consciously believing that one is a brilliant mathematician to acquire that belief, without having any interest at all in actually being a brilliant mathematician in the real world.

Now that we have a slightly clearer description of the motivational set of a subject who has a self-focused desire to believe that p , or more generally an interest in believing that p , without intrinsically desiring that p be true, or having an interest in p being true, let us ask whether the motivational set of the self-deceiver is different in any relevant respect from that of such a subject. Let us consider the case of Nicole, described by Funkhouser (2005, p. 302). Nicole is Tony's wife. She sincerely believes that she is convinced that her husband is not having an affair with her best friend, Rachel, despite having excellent evidence of the affair being at least likely. For example, Nicole's friends say that Tony's car is parked in front of Rachel's house at times when he had told Nicole he was going out with his friends; also, Tony has a significantly diminished sexual interest in Nicole; and so on. The way she reaches this conviction is instructive: it is not simply that she misinterprets the evidence that this affair may at least be likely by looking for stories to explain why such evidence should not count as conclusive; she also carefully engages in avoidance behavior, such as keeping away from Rachel's house at times when Tony says he is out with his friends, even at the cost of changing the route she would otherwise have taken. Funkhouser is in part arguing that Nicole's avoidance behavior shows that she "deep down" knows the truth, and does have the belief that her husband *is* having the affair, but I will not discuss the tenability of this here, although I have my doubts that we really need to postulate a "deep down" knowledge of how things stand in self-deception (I address this concern at length in another project and I shall briefly return to it in the next paragraph); nor am I interested in the claim that Nicole falsely believes that she believes that her husband is not having an affair, although she does not actually have the corresponding first-order belief. Rather, what I am interested in here is Funkhouser's claim that what the self-deceiver wants, or is interested in, is primarily a state of mind of belief, and that her wanting that p be true can be treated as contingent upon the self-deceiver's "desire-to-believe" want. To establish this conclusion, Funkhouser seems to be drawing here a (largely undeclared) inference to the best explanation about the motivation for avoidance behaviour: since she avoids reality, the best explanation for her avoidance of reality is that she doesn't *want* to know the

truth, that she *wants* to protect herself and her favoured opinion from the impact of reality, and so she primarily wants to believe that p . That is, in order to explain her avoidance behaviour, Funkhouser attributes to Nicole the motivation provided by a self-focused desire to believe that p as primary, while the motivation provided by the world-focused desire that p be true would be, at best, contingent upon such primary motivation. An explanation of the precise nature of such contingency is not completely spelled out in the account offered by Funkhouser. Perhaps, the contingent desire that p be true is a consequence of the primary “desire-to-believe”, in that the self-deceiver must have to do with the world anyway, in order to believe that p ; or it may be the case that the desire that p be true happens to be contingently present because of other psychological coincidences as yet unspecified. I will not develop an analysis of his view on this issue, as it would lead me too far from the major purposes of this article. What I will do, instead, is to show why I think that self-deceivers cannot be moved by the want that p be true in a merely contingent way, even if it is apparent that they very much engage in avoidance behavior. After I will have done that, I will try to make a case for an alternative, positive explanation of avoidance behavior and its motivation, that will be coherent with the view I recommend. Before I begin, I would highlight the fact that, in what follows, I am deliberately going to shift away from the “*desire-centred*” terminology used by Funkhouser and Nelkin to defend their views on the self-deceptive motivation (which terminology I have already tried to expand in passing in the first part of the paper) towards a more neutral “*want (or interest)-centred*” language. The reason why I prefer this terminological expansion will be made clear in the last portion of the paper, where I will be offering a number of considerations on “twisted” self-deception, where one believes that p even if one does not desire that p be true, and on the alleged support that it gives the self-focused desire account of the self-deceptive motivation. I will say something at that point about how I think the motivation for twisted self-deception should be analyzed and how the analysis I suggest about twisted self-deception fits my general view about the motivational content of all kinds of self-deception.

Let me start with the case for claiming that it seems unlikely that the self-deceiver’s want that p be true could be merely contingent upon her want to believe that p . One of the most distinctive features of self-deception is a more or less demanding *epistemic* work on the evidence self-deceivers have or might find, which I believe is what ultimately distinguishes self-deception from other

forms of motivated irrationality. In other forms of motivated irrationality, such as “precipitate cases” of believing (Scott-Kakures, 2002, p. 587), phenomena of jumping to conclusions under the influence of strong emotions², wishful thinking, etc., the subject involved does not normally spend time and energy on elaborating “covering stories” to justify to herself the opinion that p that she favours and does not typically struggle epistemically with evidence against p to arrive at an explanation of why it should not count as undermining p . On the contrary, self-deceivers are the champions of a pressing inner dialogue setting out to assess the strength of the evidence undermining p and balance it with the strength of the evidence suggesting that p may be true. Often they also need to share their “findings” on the strength of the evidence with others. We are all familiar with friends with documented unhappy relationships who call us on the phone at night or write us suspect letters to tell us and *explain to us why* they are really happy, *why* they are not in the position to believe that their partners are unfaithful, *why* they believe their love affair is not really over, *why* they think they have new evidence that that man or woman does like them, and so on. Most of us may, perhaps, have occasionally made such phone calls or written such letters and have experienced the inner epistemic negotiations that have encouraged us to declare our dubious conclusions to our friends. In all, self-deceivers have quite complex “convincing” stories, elaborated by means of what is generally an intense epistemic work on why they believe what they do and declare – complex stories that are lacking in other forms of motivated irrationalities and fundamentally aimed at explaining why p is true.

Now, if this epistemic work is, as I think, one of the most prominent specificities of self-deception, its fascinating and disconcerting hallmark, we may well ask if a purely self-focused want to believe that p can satisfactorily provide an explanation for the motivation that triggers such an epistemic endeavour. Certainly, the self-focused want to believe that p can sometimes successfully account for cases of artificially self-induced false beliefs or other mental states (and, with qualifications, also for other forms of motivated irrationality), but it seems fair to say that the motivation for the epistemic work that typically underpins the doxastic end-state of self-deception cannot but be a world-focused want that p be true. We need not deny that the self-deceiver is attracted to the sense of importance she attaches to believing that p , which she

² See Lazar, 1999, p. 281.

may find pleasant, or else unpleasant but important in the light of certain other wants and convictions, but her epistemic work to establish that p is true seems to show that the importance she places on believing that p is due to the importance she places on knowing that reality is such that p is true. In other words, the world-focused want that p be true is not detachable from, and contingent upon, the self-focused want to believe that p ; rather, it seems to be intrinsic to it. If it were just contingent, we would be in need of an explanation why self-deceivers try to justify their convictions epistemically. This tells us that the self-deceiver's motivational set cannot, and should not, be confused with the motivational set of someone who may even be indifferent to the way things stand and just want to acquire a state of mind for the sake of the importance to her of being in that state; nor can, or should, it be confused with the motivational set of someone who embraces a conclusion on the heat of the moment. Such confusions make us lose our grip on the nature of self-deception and seem to underdescribe the complexity of the motivation that prompts it.³

To complete the argument, we now need to try to explain why, then, self-deceivers engage in avoidance behaviour at all and how we should qualify their relationship with reality, if they are interested (as they seem) in establishing the

³ One might point out that an epistemic work could be compatible with the “desire-to-believe” account. That is, even if the subject is moved by a desire to believe, she may still need to stay focused on the world to secure her doxastic conclusion. My initial sense about this objection is that being focused on the world for the “instrumental” reason of securing a doxastic conclusion that one wants to secure is significantly different from wanting to secure a doxastic conclusion because one wants the world to be as one would like it to be. The crucial difference that I see lies fundamentally in the relationship one has with one's beliefs. If one by default relates first-personally to one's beliefs as states that are representative of the world, and does not try to manipulate those states independently of their representational goal, having with them a third-personal relationship, then, when it comes to self-deceiving, one would tend to establish the truth value of what one believes, and this is the epistemic work that I see the self-deceivers to be doing, as opposed to someone who artificially self-induces mental states. Furthermore, the whole point of the discussion is not whether someone who has a primary desire to believe can also be world-focused to secure that doxastic result – even the artificially self-inducing believer is focused; perhaps she sets out not to take another substance that can work as an antidote, within a certain time-span, and in that sense she clearly “keeps an eye” on the world. Rather, the whole discussion in hand ultimately rests on the question whether, *given what believing is*, people can really be wanting to believe that p while their wanting that p be true can be treated as merely contingent. And this is exactly why I designed cases of artificially self-inducing beliefs, as the extreme of a spectrum that helps us see that the self-deceiver more probably occupies a place close to the opposite extreme, given her intense, and presumably not merely instrumental, if she relates first-personally to her beliefs, epistemic work.

truth, or at least the credibility, of what they come to believe, but nonetheless avoid pieces of evidence that could lead them to ascertain how things actually stand.

2. Avoidance behaviour and the self-deceiver's relationship with truth and reality

Avoidance behaviour, surfacing either in self-deception or in other psychological predicaments, is undoubtedly a tricky phenomenon that may be difficult to elucidate completely, both phenomenologically and explanatorily. At the phenomenological level, it may not always be clear what is the intentional object of avoidance, if it need be consciously represented, and so on. At the explanatory level, questions arise as to *why* we engage in it at all, and what motivation there is for it. Just one thing seems to be conceptually intuitive and phenomenologically manifest: there is always something in what we avoid (an aspect of it, a thought or a feeling that it prompts, etc., however represented) that we fear or find upsetting, distressing or unacceptable. If this starting assumption is workable, no doubt those who engage in avoidance behavior ultimately do not want to get in touch with such sources of fear, upset, or distress. It is this very assumption that presumably inspired Funkhouser to conclude that self-deceivers “deep down” know the truth, but are somehow scared by it and so avoid it. On this basis, also, it is appealing to infer that *all* the self-deceivers want is to acquire the belief that *p*. I will argue in the remainder of this paragraph that while the starting assumption is correct, and it is thus likely that self-deceivers are scared by “something” that a dispassionate contact with reality may reveal, avoidance behavior does not show as yet that the leading motivation to self-deception is a “want-to-believe that *p*”, upon which the want that *p* be true would be merely contingent; on the contrary, it shows that self-deceivers have a *strong interest in p being true*, as made manifest by their epistemic work, but this *interest in p being true is heavily biased by the very want that p be true*. In order to understand this, it is crucial to unpack important details of avoidance behavior, which may easily remain undisclosed.

Let us consider again the case of Nicole. Nicole, as we have seen, avoids Rachel's house at times when Tony says he is out with his friends. An impartial observer may say that she does that ultimately because she is scared by the idea that going there might disclose to her new evidence in support of the

hypothesis that Tony is having an affair with Rachel. If such a hypothesis were confirmed, her love dreams could not survive – nor, perhaps, her marriage. So, it is manifest that, at the very least, she is scared by the way reality *may* be, that is, by the *possibility* that a certain proposition p may be true. I am clearly working here with the hypothesis testing model of self-deception made famous by Alfred Mele (2001), where self-deceivers do not start their self-deception by already believing that $\text{not-}p$ and trying to convince themselves that p ; rather, the favoured hypothesis that p is raised by the corresponding desire that p and then tested in a biased way. On the contrary, Funkhouser assumes that the self-deceiver “deep down” knows that the favoured hypothesis that p is false, and this is an additional component of his account that encouraged him to conclude that self-deceivers are not primarily motivated by world-focused desire, or, more generally, want, as appropriate: if they were, they would not try to avoid what they already know. But even if Funkhouser were right to say that they start their self-deception by already knowing how things stand, many (e.g., Bermúdez, 2000) have argued that such knowledge could have been suitably undermined by biased epistemic work on it and brought back to the status of a hypothesis in need of a new test. So, Funkhouser would have to show that in all cases of self-deception the alleged “deep down” knowledge that $\text{not-}p$ is never turned into the corresponding hypothesis that p . In this way only could he substantiate his subsequent claim that the intentional object of avoidance is the known truth, as opposed to just the possibility that a feared hypothesis may be true, but this is far from being proven by the examples he gives as they stand.

However, if the possibility that a *feared hypothesis* may be true is the intentional object of avoidance behavior, new light is shed on Nicole’s motivation to engage in such avoidance, and the tenability of the general account Funkhouser promotes about the self-deceptive motivation is deeper in trouble. For on this alternative, positive account of Nicole’s avoidance behavior, she would not be trying to avoid the truth she somehow knows as such because she is just interested in acquiring or maintaining the belief that p , but precisely *to establish that reality is the way she wants it to be*. In other words, she is so interested in establishing that p is true that she carefully avoids contact with sources of evidence suggesting that the favoured hypothesis that p may be false. In all, I acknowledge that a fatal bias affects the self-deceiver’s relationship with her interest in truth and the way the world is, but I reject the claim that the self-deceiver is only contingently moved by a concern with how she takes the world to be. The bias affecting her relationship with reality is due

to the want that p be true, in turn presumably motivated, as we will see, by other convictions, values, character traits and so on, which would have to be uncovered and described case by case. If one pursues this line of reasoning, one is led to suppose that it is the self-deceiver's want that p be true that leads her to forge the belief that p , and so, also, possibly her want to believe that p , as opposed to thinking that the want to believe that p is what triggers the whole motivational process of self-deception. On this reading of the motivation for self-deception, avoidance behaviour is acknowledged as one of the major symptoms of the fact that the self-deceiver wants the world to be such that p be true, perhaps because she fears the possibility that the world be such that p be false, which would be distressing, or discomforting, or upsetting to her. That is, avoidance behavior would be no decisive evidence for the "want-to-believe" account.

If my claim is correct, I am, however, left with the task of explaining how my favoured view could accommodate the alleged cases of "apathetic" or "indifferent" self-deception introduced by Funkhouser, and also how it deals with cases of twisted self-deception. Examples of apathetic or indifferent self-deception, according to Funkhouser, would be cases of beliefs typically acquired upon peer pressure: some people may want to believe what their peers believe without having any preference whatsoever about the truth of those beliefs. Here again, I believe such cases would need to be fully unpacked before issuing claims as to their nature. To begin with, even if we can agree that those who self-deceive upon peer pressure are attracted to the importance they attach to belonging to a group and thus sharing opinions on sensitive matters with their peers, if an epistemic work is performed by the self-deceiver to justify what she comes to believe, then we have a clue that her self-deception is not "indifferent". Secondly, the reasons why she wishes to share those opinions with her peers should be more clearly analysed. For it may be that she delegates to peers the authority to entertain true opinions on sensitive matters. On this hypothesis, she would wish to share her opinions because she takes those opinions to be true, even if she does not embark on the epistemic work to establish that p is true. Once again, the self-deception would not be "indifferent" at all. Pending further analysis, cases of apathetic self-deception should not be treated as clear cases in which the self-deceiver is primarily led to the self-deceptive belief that p by being moved by a mere want to believe that p , and with the manifest absence of any want that p be true.

3. On the significance and the prospects of unifying the leading motivation for “straight” and “twisted” self-deception

Finally, let me devote a few words to Funkhouser’s attempt at providing a unifying account of the motivation prompting two varieties of self-deception, the so called “straight” and “twisted” self-deception. According to Funkhouser, world-focused desire accounts of self-deception cannot provide a unified explanation of why some self-deceivers end up falsely believing that p while they want p to be the case (“straight” self-deception, e.g., Nicole self-deceptively believes that her husband is not having an extramarital affair while she wants him not to be having one), and others end up falsely believing that p while they do not want p to be the case (“twisted” self-deception, e.g., John self-deceptively believes that his wife is having an extramarital affair although he does not want her to be having one). That is, on the world-focused desire accounts, straight and twisted self-deception would be accounted for by two different sorts of motivations: a desire that p be the case would motivate straight self-deception, while a hostility toward p being the case would, *mysteriously*, motivate twisted self-deception. Funkhouser’s conviction is that self-focused desire accounts have the advantage of offering a unified treatment of the motivation prompting both varieties of self-deception, while explaining away the mystery affecting the motivational drive to twisted self-deception: both would be triggered by a desire to acquire a belief that p , and not by two different sorts of motivation, namely, a desire that p for straight self-deception, and a fear, or dislike, or repugnance that p for twisted self-deception. Also, besides the advantage of achieving explanatory unification across different varieties of self-deception, twisted self-deception is used in this line of reasoning as the crucial case that seems to lend the best support to the “desire-to-believe” account, as it is the kind of self-deception in which one seems to best appreciate how a desire to believe that p can move someone to believe that p , any desire that p be true clearly being absent, as the twisted self-deceiver does not desire that p be true at all. In the space available, I will just briefly set out two main clusters of considerations, largely incomplete, to provide a general outline of a line of research on this issue that I would like to develop fully elsewhere.

First, even before assessing whether twisted self-deception can be really moved by a primary want to believe that p (upon which any sort of world-focused want that p be true would be merely contingent, if not absent), it is

worth asking what explanatory advantage is gained by the explanatory unification of the two varieties of self-deception. Suppose, for the sake of argument, that the self-focused want account is correct and all sorts of self-deceivers primarily want to acquire a certain false belief. At this point, a crucial question regarding the self-deceivers of both sorts needs to be answered: why is it that straight self-deceivers want to acquire a belief that they like, while twisted self-deceivers want to acquire a belief that they dislike? I take this to be a perfectly legitimate question that all accounts of the motivation triggering self-deception should answer. A clue to the answer to this question seems to lie in the self-deceiver's relationship with reality, a relationship shaped by her values, desires, fears, other beliefs, and so on. Perhaps, the straight self-deceiver Nicole may like the belief that Tony is not having an affair because she conceives of a good marriage as based on fidelity and forcibly wants her private world to achieve this ideal; perhaps, the twisted self-deceiver John may dislike the belief that his wife is having an affair because he conceives of a good marriage as based on fidelity, but he also has a paranoid conviction that many marriages do not achieve this ideal, and so wants to test whether his is among them. Perhaps he also wants to prove to himself that his paranoid conviction is right, maybe because having a confirmation of his convictions will help him to reduce anxiety, in ways as yet unspecified. Note at this point that the answers to this question on both forms of self-deception cannot but be disjoint: different values and personalities, a different relationship with reality, as well as different fundamental wants shape the two forms of self-deception. That is, if one presses questions upon why either sort of self-deceivers is moved to self-deceive in the specific way they do, the analysis initially provided by the defender of the desire-to-believe account needs to "go deeper", in search of the deeper wants that move and shape the specific variety of self-deception in question, which a "want-to-believe" account does not seem to trace. In other words, the explanatory unification seems to be achieved at the expense of an in-depth grasp of the individual's specific motivation to achieve a particular kind of self-deception.

The stage is perhaps now better set to turn to the second cluster of thoughts I would like to promote. I would like now to consider briefly the very nature of motivation that drives twisted self-deception as such, independently of the prospects of the explanatory unification I have discussed in brief. My intention was to foreshadow my fundamental instinct towards the issue of what may move someone to self-deceive in believing what she fears, or finds

upsetting, or anyway does not like to be so, earlier in the paper by deliberately shifting, on as many occasions as I could, away from the “desire-centred” analysis of the various self-deceptive motivations, including twisted self-deception, and the related terminology, towards a broader “want-centred” analysis of them. When I think of cases of twisted self-deception, and when I think of cases of straight self-deception as well, two general features of the motivational set of a self-deceiver of any sort strike me as central: 1) it is not clear that a desire for something is invariably the most crucial, or deepest, motivational drive for a subject, in general, both in the practical domain, where we typically analyse motivation for action, decisions and so on, and in the theoretical domain, where we can sometimes, as in the case of self-deception, trace motivation to a certain reasoning, direction of cognition, and so on; 2) even when a desire presents itself as the motivational drive for an action or a train of thought etc., it seems to me that the desire-driven motivation is not necessarily the whole motivation story that one could in any case tell to explain the case in point. The analysis of a motivational set can very often, though perhaps not always, “go deeper”, and it should do so, if appropriate. I believe that many, even if presumably not all, cases of both straight and twisted self-deception hold out material for deeper analysis, which may prove to be instructive as to the tenability of any “desire-to-believe” account of the self-deception motivation, and more generally, of any “want-to-believe” account. It seems to me likely that deeper drives, further motivations, typically shaped by ground values and more or less hidden convictions of various kinds that a subject has, forge the “surface-motivation” for self-deception, which can be either a desire, or a fear, or any other motivational state one may find appropriate to attribute to a subject in a specific case. “Deeper down”, however, there seems to be a much more complex psychological world to explore. As far as I can see, it is this exploration only, subtle and demanding as it may be, that can allow us to hope for a chance of grasping the specific motivation that ultimately moves someone to engage in a specific sort of self-deception. General human drives may be identified in all sorts of self-deception, of course, but while many think that one of the most ever-present motivational drives attributable to a subject who self-deceives is a desire for something, I have the feeling that a deep fear might instead better explain what the subject wants and seeks by self-deceiving – fear of the psychological pain that certain possible states of affairs may cause. If one believe that there is no renouncing the desire-centred analysis, it is perhaps more tempting to look for

an overarching desire – in the account in question, a “desire to believe” – which is then thought to be capable of unifying both straight and twisted self-deception, which unquestionably seem to be moved, but perhaps only superficially, by a desire and a fear respectively.

So, to sum up the general thoughts that guide my research on the issue:

- a) The deeper motivational drive for self-deception may well be a fear, instead of a desire – as I said, fear of the psychological pain that certain possible states of affairs may cause, to be coupled with the psychological specificities of the subject involved, her other beliefs and values, her other wants, which only can explain why her self-deception was triggered and what she ultimately want as a person, more than simply as a self-deceiver. Self-deception is an extraordinary window on the psychological structure of an individual subject, and I believe that this explanatory richness should not be lost for the sake of any unification, still less for a unification in the name of desires⁴.
- b) Once one takes this route, and does not look for a desire only (overarching or otherwise) to explain self-deception in general, but rather looks for deeper wants, case by case, twisted self-deception, in which desires are not (at least superficially) prominent, takes on a new light and seems to have a chance of being accommodated, in ways as yet unspecified, in a “want-centred” account.
- c) It also seems easy to accommodate in a “*world-focused want*” account. Twisted self-deception has been considered as the variety of self-deception that lends best support to the “desire-to-believe” account, given that there is no manifest world-focused desire that p , but rather a fear that p , so if a desire is thought to be motivationally necessary to move self-deception, and no world-focused desire is present in twisted self-deception, then having a “self-focused desire-to-believe account” at hand may seem helpful. But if twisted self-deception is shown to be

⁴ There might be a worry of regress, here, about the ultimate motivational source. It may be said that the fundamental fear itself may be due to a desire to feel pleasure. The issue is intriguing, and to deal with it satisfactorily would take me too far from the present purposes. For the time being, I just remark that the drive towards pleasure, and the desires that spring from it, and the fear of feeling pain, and the more specific fears it causes, might well be the two sides of one and the same coin, and so one depends upon the other; yet, what it is most salient, and causally primary, in one specific case of self-deception as opposed to another, may still be one side only of the coin.

driven by other wants, perhaps driven in turn by the want not to feel pain caused by specific contact with a specific reality, thematically sensitive for an individual, then twisted self-deception begins to appear less problematic for a “world-focused want” account than might initially have been thought.

- d) Finally, if straight self-deception is equally deeply driven not by superficial wants, but rather by wants that go deeper than the surface desire that p be true, maybe an explanatory unification can still be achieved, although on different grounds.

I hope that I have, in the space available, at least established the general theoretical background for a future project I wish to pursue about the motivation for self-deception, and that I have sufficiently clearly set out my reasons for exploring views towards which my philosophical instinct tends to lead me.

REFERENCES

- Bermúdez, J.L. (2000). Self-Deception, Intentions and Contradictory Beliefs. *Analysis*, 60(4), 309–319.
- Funkhouser, E. (2005). Do the Self-Deceived Get What They Want?. *Pacific Philosophical Quarterly*, 86(3), 295–312.
- Lazar, A. (1999). Deceiving Oneself or Self-Deceived? On the Formation of Belief “Under the Influence”. *Mind*, 108(430), 265–290.
- Mele, A. (2001). *Self-Deception Unmasked*. Princeton: Princeton University Press.
- Nelkin, D.K. (2002). Self-Deception, Motivation, and the Desire to Believe. *Pacific Philosophical Quarterly*, 83(4), 384–406.
- Scott-Kakures, D. (2002). At “Permanent Risk”: Reasoning and Self-Knowledge in Self-Deception. *Philosophy and Phenomenological Research*, LXV(3), 576–603.

