

RESPONSIBILITY FOR SELF-DECEPTION

Marie van Loon

Volume 13, numéro 2, été 2018

URI : <https://id.erudit.org/iderudit/1059502ar>

DOI : <https://doi.org/10.7202/1059502ar>

[Aller au sommaire du numéro](#)

Éditeur(s)

Centre de recherche en éthique (CRÉ)

ISSN

1718-9977 (numérique)

[Découvrir la revue](#)

Citer cet article

van Loon, M. (2018). RESPONSIBILITY FOR SELF-DECEPTION. *Les ateliers de l'éthique / The Ethics Forum*, 13(2), 119–134. <https://doi.org/10.7202/1059502ar>

Résumé de l'article

Dans cet article, je soutiens que la conception d'auto-illusion chez Alfred Mele remplit toujours l'une des conditions de la responsabilité doxastique, à savoir la « sensibilité aux raisons » (reasons-responsiveness). Il en est ainsi car les mécanismes d'auto-illusion dans la formation de croyances produisent des types de croyances qui remplissent les critères pour la responsabilité doxastique quant à la sensibilité aux raisons épistémiques. J'explique pourquoi dans cet article. Mele décrit la relation du sujet à la preuve comme biaisée. Le sujet ne croit pas seulement sur la base de preuves, mais de preuves manipulées. Mele avance quatre façons qu'a le sujet de faire ceci. Le sujet peut mal interpréter positivement ou négativement, focaliser de façon sélective, ou accumuler des preuves. Par ces formes de manipulations, la preuve est formulée de sorte qu'elle produise un fondement pour la croyance en une proposition qui s'accorde avec le désir du sujet que P. Peu importe la forme de manipulation qu'emploie le sujet, la preuve contre P doit être neutralisée d'une façon ou d'une autre. Une neutralisation réussie de la preuve requiert la capacité de reconnaître ce que soutient la preuve et la capacité d'y réagir. Ces capacités consistent précisément en ces deux parties de la condition de la sensibilité aux raisons, soit la réceptivité et la réactivité aux raisons. En ce sens, les croyances d'auto-illusion remplissent toujours la condition de la sensibilité aux raisons pour la responsabilité doxastique. Toutefois, étant donné que la sensibilité aux raisons n'est une condition nécessaire que pour la responsabilité doxastique, cela ne veut pas dire que les sujets souffrant d'auto-illusion sont toujours responsables de leurs croyances.



RESPONSIBILITY FOR SELF-DECEPTION

MARIE VAN LOON

PHD STUDENT, UNIVERSITY OF ZURICH

ABSTRACT:

In this paper, I argue that Alfred Mele's conception of self-deception is such that it always fulfils the reasons-responsiveness condition for doxastic responsibility. This is because self-deceptive mechanisms of belief formation are such that the kind of beliefs they bring about are the kind of beliefs that fulfil the criteria for doxastic responsibility from epistemic reasons responsiveness. I explain why in this paper. Mele describes the relation of the subject to the evidence as a biased relation. The subject does not simply believe on the basis of evidence, but on the basis of manipulated evidence. Mele puts forward four ways in which the subject does this. The subject could misinterpret positively or negatively, selectively focus, or gather evidence. Through these ways of manipulation, the evidence is framed such that the final product constitutes evidence on the basis of which the subject may believe a proposition that fits that subject's desire that P. Whichever form of manipulation the subject uses, the evidence against P must be neutralized in one way or another. Successful neutralization of the evidence requires the ability to recognize what the evidence supports and the ability to react to it. These abilities consist precisely in the two parts of the reasons-responsiveness condition, reasons receptivity and reasons reactivity. In that sense, self-deceptive beliefs always fulfil the reasons-responsiveness condition for doxastic responsibility. However, given that reasons responsiveness is only a necessary condition for doxastic responsibility, this does not mean that self-deceived subjects are always responsible for their belief.

RÉSUMÉ :

Dans cet article, je soutiens que la conception d'auto-illusion chez Alfred Mele remplit toujours l'une des conditions de la responsabilité doxastique, à savoir la « sensibilité aux raisons » (reasons-responsiveness). Il en est ainsi car les mécanismes d'auto-illusion dans la formation de croyances produisent des types de croyances qui remplissent les critères pour la responsabilité doxastique quant à la sensibilité aux raisons épistémiques. J'explique pourquoi dans cet article. Mele décrit la relation du sujet à la preuve comme biaisée. Le sujet ne croit pas seulement sur la base de preuves, mais de preuves manipulées. Mele avance quatre façons qu'a le sujet de faire ceci. Le sujet peut mal interpréter positivement ou négativement, focaliser de façon sélective, ou accumuler des preuves. Par ces formes de manipulations, la preuve est formulée de sorte qu'elle produise un fondement pour la croyance en une proposition qui s'accorde avec le désir du sujet que P. Peu importe la forme de manipulation qu'emploie le sujet, la preuve contre P doit être neutralisée d'une façon ou d'une autre. Une neutralisation réussie de la preuve requiert la capacité de reconnaître ce que soutient la preuve et la capacité d'y réagir. Ces capacités consistent précisément en ces deux parties de la condition de la sensibilité aux raisons, soit la réceptivité et la réactivité aux raisons. En ce sens, les croyances d'auto-illusion remplissent toujours la condition de la sensibilité aux raisons pour la responsabilité doxastique. Toutefois, étant donné que la sensibilité aux raisons n'est une condition nécessaire que pour la responsabilité doxastique, cela ne veut pas dire que les sujets souffrant d'auto-illusion sont toujours responsables de leurs croyances.

I. INTRODUCTION

In the same way that we reproach the ignorant subject, “You should have known,” we reproach the self-deceiver, “You shouldn’t be self-deceived.” Both the ignorant subject and the self-deceived subject make an *epistemic*¹ mistake, for which they are to blame at least epistemically and sometimes morally. Since blame requires responsibility, holding self-deceivers responsible for their self-deceptive belief must be appropriate. If these assumptions are correct, then there should be a way of explaining why self-deceptive beliefs are beliefs for which we are responsible.

According to Alfred Mele, self-deception need not be conceived on the model of interpersonal deception but may simply be understood as a motivationally biased belief (Mele, 1997; 1999; 2001; 2006). Because, according to such a conception, self-deception consists in, among other things, holding a belief, a theory of doxastic responsibility should be able to explain how we can be responsible for self-deceptive beliefs. Among the various competing theories of doxastic responsibility, *reasons responsiveness* (McHugh, 2013; 2014; 2015) offers a necessary condition for doxastic responsibility.

In this paper, I argue that Mele’s conception of self-deception is such that it always fulfils the reasons-responsiveness condition for doxastic responsibility. This is because self-deceptive mechanisms of belief formation are such that the kind of beliefs they bring about are the kind of beliefs that fulfil the criteria for doxastic responsibility and more particularly the criteria for doxastic responsibility from epistemic reasons responsiveness. I explain why in this paper. Mele describes the relation of the subject to the evidence as a biased relation. The subject does not simply believe on the basis of evidence, but on the basis of manipulated evidence. Mele puts forward four ways in which the subject does this. The subject could misinterpret positively or negatively, selectively focus, or gather evidence. What matters is that the evidence is framed such that the final product constitutes evidence on the basis of which the subject may believe a certain proposition. This strategy requires avoiding coming into contact with data that would disprove *P*. If the subject does not manage to avoid it, then that subject might instead misinterpret such information as counting in favour of *P* or at least as not counting against *P*. Whichever form of manipulation the subject uses, the evidence against *P* must be neutralized in one way or another. Successful neutralization of the evidence requires the ability to recognize what the evidence supports and the ability to react to it.

Reasons responsiveness requires that subjects would recognize evidence against *P* if they were presented with it and that they would react to this evidence if they were presented with it. These conditions that require the subject to recognize and react to evidence in counterfactual scenarios express an ability that is expected from the subject, and it is the possession of this ability that makes a belief reasons responsive.

I argue that, in order to successfully self-deceive, an epistemic subject must possess this ability. Possessing this ability makes a belief reasons responsive. In that sense, self-deceptive beliefs always fulfil the reasons-responsiveness condition for doxastic responsibility. However, given that reasons responsiveness is only a necessary condition for doxastic responsibility, this does not mean that self-deceived subjects are always responsible for their belief.

In order to show this, I proceed in the following way. In section II, I summarize the reasons-responsiveness account of doxastic responsibility. In section III, I offer a reminder of Mele's theory of self-deception. In sections IV, V, and VI, I apply reasons responsiveness to self-deception.

II. EPISTEMIC REASONS RESPONSIVENESS

According to the reasons-responsiveness account of doxastic responsibility defended by Conor McHugh (2013; 2014; 2017), subjects are responsible for their beliefs and other attitudes in virtue of a feature they must possess: reasons responsiveness.

I speak of doxastic responsibility here because I take it that the object under consideration is a belief and that we cannot assess the responsibility for self-deception without using an adequate theory that pertains to its particular object. Most authors take on assessing responsibility for self-deception by asking whether subjects are morally responsible for it. I differ in that I first want to ground moral responsibility in doxastic responsibility.

One dominant account of doxastic responsibility is McHugh's application of Fischer and Ravizza's reasons-responsiveness account (Fischer and Ravizza, 1998) to the epistemic case, epistemic reasons responsiveness. Whereas Fischer and Ravizza's reasons responsiveness provides necessary conditions for responsibility for actions, McHugh provides us with necessary conditions for beliefs. The crucial difference is that, in the epistemic version of reasons responsiveness, we are not after any reasons that agents would recognize for holding their belief and that constitute the basis on which they form, maintain, or revise their belief, but we are specifically after epistemic reasons—that is, the evidence that agents would recognize for holding their belief and that constitute the basis on which they form, maintain, or revise their belief. The difference is important because practical reasons and epistemic reasons do not function in the same way with regard to belief. It is indeed frequently admitted that we cannot believe a proposition for practical reasons. Thus, the only reasons that are pertinent to the responsiveness of belief are epistemic reasons. Therefore, there is an important difference in applying epistemic reasons responsiveness to self-deception rather than simple reasons responsiveness to it—not only because these theories are different but also because of the reason why there is an *epistemic* reasons responsiveness in the first place: we cannot ask agents to be responsive to nonepistemic reasons in order for them to be responsible for their belief.

Levy (2004) and DeWeese-Boyd (2007) and Nelkin (2013), who discuss the application of reasons-responsiveness accounts of moral responsibility to self-deception, seem rather concerned with the actions and omissions (mental and nonmental) that lead to or maintain a state of self-deception.² In this vein, they are concerned with responsibility for self-deception in virtue of the indirect control we have over these beliefs. Such indirect control operates through actions (and omissions). Given that these authors are asking whether we can be held responsible for self-deception in virtue of these actions and omissions that surround self-deception, it makes perfect sense to turn to the Fischer and Ravizza version of reasons responsiveness, since they, too are interested in responsibility for actions. I am concerned not with questions of responsibility for the actions that lead to and follow from being in a state of self-deception, but with the state itself—that is, the self-deceptive belief. For this reason, McHugh’s reasons responsiveness, since it concerns doxastic responsibility directly, is better fitted for the task of asking whether and why subjects may be held responsible for their self-deceptive belief.

Theories of doxastic responsibility try to explain why there can be responsibility for beliefs despite doxastic involuntarism. Doxastic involuntarism is the thesis that beliefs are not under our *direct control* (Williams, 1973; Alston, 1988). We may work to abandon them or form new ones by means of other actions or attitudes, but not in the same way that we are able to raise an arm at will. Because we lack direct control over beliefs, it may seem as if we can never be held responsible for our beliefs. Several solutions to this problem have been offered in the literature. Most³ theories of doxastic responsibility aim at grounding responsibility for beliefs on a basis that does not involve direct control. Some argue that what is needed is a form of indirect control (Meylan, 2013; 2017; Peels, 2013; 2017), others argue that we are responsible for beliefs if they are intentional (Steup, 2012), and others explain that we are responsible for beliefs in virtue of the fact that beliefs are the kind of attitude that reflect our take on the world (Hieronymi, 2008).

Reasons responsiveness remains the most agnostic of these theories with regard to the exact nature of control over beliefs; it endorses neither direct nor indirect control. In this sense, the type of control reasons responsiveness requires for one to be responsible for one’s beliefs is minimal. Mele’s conception of self-deception as motivationally biased belief intuitively doesn’t seem to require any kind of doxastic control that is not minimal, whether direct or indirect. Given that reasons responsiveness requires only a minimal form of control, this makes it perfectly well suited to accommodate Mele’s model of self-deception.

McHugh proposes a form of control that is distinct from direct control and indirect control and which applies specifically to beliefs and other attitudes: *attitudinal control* (McHugh, 2017). Nevertheless, this particular notion of control is a very minimal one, more akin to a kind of sensitivity to evidence. *S* has attitudinal control over *S*’s belief that *P* only if⁴ *S*’s belief is responsive to reasons. *S*’s belief that *P* is responsive to reasons only if both of the following conditions hold.

Reasons Receptivity

S would recognize epistemic reasons⁵ to believe *P* if presented with such reasons to believe *P*.

Reasons Reactivity

S would react to epistemic reasons *S* has to believe *P*.

Reasons receptivity and reasons reactivity together form epistemic reasons responsiveness (McHugh, 2013; 2014; 2015). Reasons responsiveness is a necessary condition for doxastic responsibility.

A belief fulfils the reasons-receptivity condition if in a sufficiently wide range of counterfactual scenarios relevant to the actual scenario the subject recognizes evidence counting in favour of or against *P*. Being reasons receptive amounts to being able to recognize that some evidence would count in favour of or against *P* if one were presented with such evidence. For example, suppose I believe that the teapot on my desk contains Earl Grey. My belief is reasons receptive if, were there Lapsang Souchong in the teapot, I would recognize the smoky scent emanating from the teapot as evidence counting against the belief that the teapot is full of Earl Grey.

A belief fulfils the reasons-reactivity condition if the subject, when faced with evidence to the contrary, reacts to that evidence by revising the belief, and would do so in a sufficiently wide range of counterfactual scenarios relevant to the actual scenario. For example, take again my belief that the teapot on my desk contains Earl Grey. Given that the smoky scent emanating from the teapot would be evidence supporting the belief that the teapot contains Lapsang Souchong and not Earl Grey, I am reactive to evidence if I abandon the belief that the teapot contains Earl Grey in reaction to this new piece of evidence.

Beliefs that are not reasons responsive are beliefs for which we are not responsible—beliefs involved in paranoia, for example (McHugh, 2014). In these cases, the belief fails to be reasons receptive. The subject would not recognize evidence supporting not-*P* if he or she were presented with such evidence in a wide range of counterfactual scenarios relevant to the actual scenario. For example, take the true belief that there is someone sitting behind me in the café. If in a wide range of relevant counterfactual scenarios I turn around, see no one at the table behind me, and yet do not recognize this as evidence that there is no one sitting behind me, I am not receptive to evidence. This is one way to fail to be reasons responsive. Another way to fail to be reasons responsive is to fail to be reactive to reasons. For McHugh, what he calls a “repressed belief” typically fails to react to recognized evidence. By “repressed beliefs,” McHugh means implicit prejudiced beliefs, for example (McHugh, 2017; p. 2752). In spite of having the ability to recognize reasons against their belief, subjects are not able to revise it. Here the subjects are not reasons reactive, and thus not reasons responsive. Thus, on the reasons-responsiveness account of doxastic responsibility, the paranoid subjects and subjects with repressed beliefs are not responsible for their beliefs.

In the next section, I turn my attention to self-deception itself and more precisely to Alfred Mele's account, of which I provide an overview.

III. MELEAN SELF-DECEPTION

The following description of self-deception is generally agreed upon in the literature. A subject is self-deceived in believing a certain proposition when that subject seems to persist in believing this proposition in spite of the evidence he or she has against this proposition. In the majority of cases present in the literature, self-deceptive beliefs are emotionally significant for the subject: beliefs about the faithfulness of partners, beliefs about the morality of your loved ones, beliefs about our own value or emotional states, and the like. Mele proposes the following set of jointly sufficient conditions for self-deception (Mele, 1997; 1999; 2001; 2006):

- 1) The belief that p which S acquires is false.
- 2) S treats data relevant, or at least seemingly relevant, to the truth value of p in a motivationally biased way.
- 3) This biased treatment is a non-deviant cause of S 's acquiring the belief that p .
- 4) The body of data possessed by S at the time provides greater warrant for $\neg p$ than for p .

As will become apparent, if reasons responsiveness is to be found in self-deception, its locus must be situated in the way the self-deceived subject treats data or evidence: indeed, in explaining doxastic responsibility, reasons responsiveness focuses on the way in which the subject recognizes and reacts to evidence. For this reason, I focus on Mele's second condition for self-deception and the way in which one treats the evidence.

Note that Mele mentions no desire in his conditions. The subject's desire that P , however, seems to be implicit in the second condition, which states that the subject treats the evidence in a *motivationally* biased way. For the present purpose, I will take the subject's motivation to be a desire. "Biased" picks out the manipulative strategies at work in the formation or sustaining of the subject's beliefs. Take the following example.

Annie and Alvy

Annie wishes she were still in love with Alvy. Because of her desire, she interprets the frequent knots in her stomach in his proximity as evidence that she is still in love—whilst in reality these knots are evidence of a growing anxiety in his presence—and forms the belief that she is still in love with Alvy. She also ignores the fact that she has lost interest in what Alvy does and frequently avoids physical contact. Annie is self-deceived in believing that she is still in love with Alvy.

Here Annie (1) forms the false belief that she is still in love with Alvy; (2) treats the evidence relevant to whether she is still being in love with Alvy in a motivationally biased way, by interpreting, because of her desire to still be in love with him, the knots in her stomach as evidence of her still being in love with him; (3) treats the evidence in such a way that it nondeviantly causes her to form the belief that she is still in love with Alvy; and (4) possesses a body of evidence that provides greater warrant for believing that she is not in love with Alvy anymore than for believing that she is still in love with Alvy.

Condition (2) requires that the self-deceived subject treat evidence in a motivationally biased way. This means that the subject acquires a belief that has been formed thanks to some biasing strategies. The biasing strategies in play may result in, but are not restricted to, the following effects: negative misinterpretation of the evidence, positive misinterpretation of the evidence, selective focusing on the evidence, and selective evidence gathering (Mele, 2001). In what follows, I provide the details of these ways of manipulating the evidence and see how these might play out in “Annie and Alvy.”

In cases of negative misinterpretation, the subject does not count as supporting not-*P* (where *P* stands for “Annie is still in love with Alvy”) evidence that would be recognized as supporting not-*P* in the absence of her desire that *P* (Mele, 2007). In “Annie and Alvy,” there must be evidence that Annie would usually count as supporting not-*P* but that she does not count, in this scenario, as supporting not-*P*—e.g., that her heart does not race when Alvy is nearby.

In cases of positive misinterpretation, the subject counts as supporting *P* evidence that would be recognized as supporting not-*P* in the absence of his or her desire that *P* (Mele, 2007). Again, in “Annie and Alvy,” there could also be evidence that Annie would usually count as supporting not-*P* but that she counts, in this scenario, as supporting *P*—e.g., the knots in her stomach.

In cases of selective focusing, the subject fails to focus on evidence that that subject would usually count as supporting not-*P* in the absence of his or her desire and focuses on evidence that seems to support *P* (Mele, 2007). Recurring panic attacks in the presence of Alvy would be an example of evidence that Annie would usually count as supporting not-*P* in normal circumstances but that she ignores in the actual scenario, where she is self-deceived. Instead she focuses on evidence that seems to support *P*—e.g., that she enjoyed Alvy’s last steamed lobster.

In cases of selective evidence gathering, the subject overlooks evidence supporting not-*P* that would have been easy to obtain in the absence of his or her desire that *P* and finds evidence seemingly supporting *P* that would have been hard to find in the absence of his or her desire that *P*. For example, Annie might avoid opening her journal and instead spend time looking at pictures of their last holiday together. The crucial difference with selective focusing on evidence is that, whereas the latter consists in cognitive operations (e.g., remembering, forgetting,

ignoring, etc.), the former seems to consist in concrete actions (e.g., searching, collecting, looking, reading or not reading one's journal, etc.).

Negative misinterpretation, positive misinterpretation, selective focusing, and selective evidence gathering all consist in successfully manipulating the evidence in such a way that it fits the subject's desire that *P*.

Mele's account captures perfectly well a feature of the phenomenon I have described at the beginning of this section. This is the feature of self-deception that consists in the subject's being in touch with the evidence in one way or another. The subject seems indeed to entertain a paradoxical relation with what that subject believes to be the facts. As I have shown in my presentation of reasons responsiveness, the subject's relation to the evidence is the locus of doxastic responsibility. Indeed, for a subject to be considered responsible for his or her belief, that belief must be reasons responsive. Another way of saying this is that subjects may be held responsible for their belief only if they follow the norms that govern belief formation, maintenance, and revision. According to these norms, subjects should be able to recognize (the right kind of) reasons for their belief if presented with them and should be able to react to these same reasons. This norm concerns the subject's relation to evidence. Thus, if we are to assess the reasons responsiveness of self-deception—to assess whether subjects fulfil a necessary condition for being held responsible for their self-deceptive beliefs—we will have to take a closer look at the feature of self-deception that pertains to the subjects' relation to the evidence. This is what I do in the next section. If it turns out that in cases of self-deception a subject would indeed recognize and react to epistemic reasons in a wide and relevant range of counterfactual scenarios, then the subject's belief is reasons responsive and moreover fulfils a necessary condition for doxastic responsibility.

IV. ACTUAL-SEQUENCE MECHANISM

Before I go on to ask whether self-deceptive beliefs fulfil both reasons receptivity and reasons reactivity, it is important to clarify the range of counterfactual scenarios that must be examined in order to determine whether self-deceptive beliefs meet the reasons-responsiveness condition. Delimiting the relevant range of counterfactual scenarios depends on the mechanism of belief formation in the actual sequence, also known as *the actual-sequence mechanism*. In this section, I identify this mechanism as the manipulation of the evidence that takes place in self-deception and that enables self-deceived subjects to form a belief that fits their desire.

What determines the relevance of the counterfactual scenario is the mechanism of belief formation used in the actual scenario, *the actual-sequence mechanism*. The mechanism must be the same across a wide range of counterfactual scenarios. It is however unclear what the conditions for the individuation of the actual-sequence mechanism are (Ginet, 2006; McKenna, 2013), and defining them remains to be done. This is a problem not only for McHugh's epistemic reasons

responsiveness, but for Fischer and Ravizza's original reasons-responsiveness account, too, from which he borrows.

Fischer and Ravizza themselves provide the following explanation:

We must confess that we do not have any general way of specifying when two kinds of mechanisms are the same. This is a potential problem for our approach; it will have to be considered carefully by the reader. But rather than attempting to say much by way of giving an account of mechanism individuation, we shall simply rely on the fact that people have intuitions about fairly clear cases of "same kind of mechanism" and "different kind of mechanism". For example, we rely on the intuitive judgement that the normal mechanism of practical reasoning is different from deliberations that are induced by significant direct electronic manipulation of the brain, hypnosis, subliminal advertising, and so forth. (Fischer and Ravizza, 1998, p. 40)

The individuation of the actual-sequence mechanism is a general difficulty for reasons responsiveness, and my concern is not to solve it here. Instead I will rely on McHugh's own indication regarding the epistemic case, that "the actual sequence mechanism must be owned by the agent" and "might be things like perception, memory and reasoning" (McHugh, 2013, p. 143).

One option is to narrow the mechanism down to these simple mechanisms—i.e., perception, memory, reasoning, etc.—by excluding the subject's desire and biases from the mechanism. The problem with this option is that, if the relevant counterfactual scenarios are those in which the subject is not motivationally biased, then the relevant counterfactual scenarios are scenarios where the subject is not self-deceived in the Melean sense. This is rather odd and gives the impression that we wouldn't be assessing the reasons responsiveness of self-deceptive beliefs only, but those of non-self-deceptive beliefs as well.

It might also be that the actual sequence mechanism consists in the interpretation of evidence in a motivationally biased way—in short, that it consists in the manipulation of the evidence (i.e., negative misinterpretation, positive misinterpretation, selective focusing, and selective gathering). This claim entails that the manipulation of the evidence is a mechanism in its own right, akin to mechanisms such as perception, memory, and reasoning. One might object that, even though manipulation might count as a type of mechanism, there seems to be an important difference between manipulation, on the one hand, and perception, memory, and reasoning on the other: contrarily to perception, memory, and reasoning, manipulation is a composite of simpler mechanisms. It might be, indeed, that manipulation of the evidence requires perceiving the evidence in the first place. However, the same holds for memory. Therefore, if McHugh identifies memory as a legitimate mechanism of belief formation, then conferring the same status onto manipulation should not be an issue. That manipulation constitutes a mechanism in its own right does not seem too unreasonable;

it is after all the way in which a subject acquires a self-deceptive belief. In the same way that a subject comes to believe by means of perceiving, the self-deceived subject comes to believe by means of manipulating the evidence. When we ask which mechanism of belief formation led the subject to believe deceitfully, it is more likely that we point at the manipulation of the evidence rather than at a noncomposite form of mechanism.

I have identified the actual-sequence mechanism in the case of self-deception with the manipulation of the evidence. What has not been determined is whether the manipulation of the evidence should remain of the very same kind (that is, fixed) throughout the counterfactual scenarios (e.g., positive misinterpretation)—this is manipulation narrowly conceived. Alternatively, the manipulation of the evidence could be allowed to adjust to different ways of becoming motivationally biased (e.g., positive misinterpretation at w_1 , negative misinterpretation at w_2 , selective focusing at w_3 , ..., w_n , where w_1 , w_2 , and w_3 , ..., w_n , are relevant counterfactual scenarios)—this is manipulation broadly conceived. I presuppose that allowing the kind of manipulation to vary across the counterfactual scenarios is analogous to allowing the way one visually perceives to vary (e.g., wears glasses at w_1 , squints at w_2 , etc.), hence the general advice for the individuation of the actual-sequence mechanism is not violated. It would make little sense not to allow the kind of manipulation to vary. After all, if the kind of manipulation does not vary depending on the epistemic reasons or on the evidence with which the subject would be confronted, the relevant range of counterfactual scenarios would be overly restricted. That is, the range would be restricted to counterfactual scenarios in which, for example, the subjects positively misinterpret their evidence in order to believe what they desire. Because it only makes sense that the evidence counting against what the subject desires would be positively misinterpreted by the subject, the range of counterfactual scenarios would also be restricted to counterfactual scenarios in which the subject is confronted with evidence counting against the subject's desire (if we're looking at belief formation) or belief (if we're looking at belief maintenance). A range of counterfactual scenarios thus restricted is in fact not relevant, in the sense that the very point of thinking about reasons responsiveness in counterfactual terms is to look at what would happen if the subject were confronted with evidence that is different than in the actual scenario. Let us add here as well that what matters when it comes to individuation of the actual-sequence mechanism is that it would clearly be the agent's or the subject's own—that acting or believing due to a certain mechanism would be distinct from acting or believing due to a mechanism that is not the agent's or subject's own, having been implanted in his or her brain by an external person, for example. For these reasons, I will assume that the ways in which the subject manipulates the evidence should be allowed to vary.

V. REASONS RECEPTIVITY

In this section, I assess the reasons receptivity of self-deceptive beliefs. For any S , where S is a self-deceived subject, to fulfil the reasons-receptivity require-

ment, the following proposition must be true: *S* is self-deceived and *S* would recognize evidence against *P* across a wide range of relevant counterfactual scenarios. In other words, it must be the case that “*S* is self-deceived” and “*S* would recognize evidence against *P* across a wide range of relevant counterfactual scenarios” can be true at the same time.

Let’s start with the actual scenario. In the actual scenario, *S* becomes self-deceived, let’s suppose, by means of positive misinterpretation. *S*’s desire that *P* leads *S* to count as supporting *P* evidence that supports not-*P*. For example, Annie’s desire to still be in love with Alvy leads her to interpret her stomach knots as support for the belief that she’s still in love, though the evidence in fact supports the belief that she’s not in love anymore (let’s say the knots are caused by her anxiety). For Annie to be receptive to reasons, there must a wide range of relevant counterfactual scenarios in which she would recognize evidence supporting not-*P* (that she’s not in love with Alvy anymore). The basic idea is that Annie must be sensitive to alternative evidence across these scenarios.

If the actual-sequence mechanism is the manipulation of the evidence, then the relevant range of counterfactual scenarios are the ones where Annie comes to believe a proposition by manipulating the evidence. In these counterfactual scenarios, evidence, whether it supports *P* in some or not-*P* in others, is manipulated and interpreted as supporting *P*. Because Annie comes to believe *P* in all the relevant counterfactual scenarios, it might look as though, at first sight, her belief is not receptive to reasons. That she does might result in Annie’s belief not being reactive to reasons, but for now let us simply look at the receptivity of her belief.

If we identify the mechanism of belief formation as the manipulation of the evidence itself, the relevant range of counterfactual scenarios are the ones where *S* comes to believe that *P* by manipulating the evidence in various ways. Thus, for each variation in the evidence, during which the desire that *P* remains constant, the way of manipulating the evidence is adjusted to the evidence at hand. In these counterfactual scenarios, *S* recognizes evidence supporting *P*, or evidence supporting not-*P* *in order to be further taken as supporting P*. In “Annie and Alvy,” the evidence across the counterfactual scenarios might vary: at *w*₁, Annie has knots in her stomach, at *w*₂ she does not, at *w*₃ she has a panic attack, etc. In each counterfactual scenario, the way of interpreting the evidence changes so as to adjust to what is required in order to satisfy Annie’s desire to still be in love with Alvy. At *w*₂, for example, Annie might resort to negative misinterpretation by not taking the absence of knots in her stomach as evidence supporting that she is not in love. Evidence at *w*₃ might call for selective focusing. The point is that such adaptation requires the ability to recognize evidence. Thus, according to a conception of the manipulation of the evidence broadly conceived, self-deceptive beliefs fulfil reasons receptivity. If they fulfil reasons reactivity as well, then self-deceptive beliefs are responsive to reasons.

VI. REASONS REACTIVITY

I now turn to reasons reactivity. For S , where S is a self-deceived subject, to fulfil the reasons-reactivity requirement, the following proposition must be true: S would react to evidence against P across a wide range of relevant counterfactual scenarios. In other words, it must be the case that “ S would react to evidence against P across a wide range of relevant counterfactual scenarios” must be true at the same time. I hinted at the fact, in the previous section, that, because the self-deceived subject forms the belief that P no matter the evidence he or she recognizes, it might look as if self-deceptive beliefs fail to meet the reasons-reactivity part of reasons responsiveness.

In the actual scenario, S becomes self-deceived, let us suppose again, by means of positive misinterpretation. S 's desire that P leads S to count as supporting P evidence that supports not- P and further to form the belief that P . Annie's desire to still be in love with Alvy would lead her to count her stomach knots as supporting the belief that she is still in love, though the evidence in fact supports the belief that she is not in love anymore (suppose the knots are caused by her anxiety), and, on this basis, she would form the belief that she is still in love. For Annie to be reactive to evidence, there must a wide range of relevant counterfactual scenarios in which she reacts to evidence she recognizes as supporting not- P by forming a belief accordingly. In fact, Annie would react to evidence she recognizes as supporting not- P by forming a belief *contrary* to what she recognizes as evidence. This way of failing to meet reasons reactivity, if it is indeed the case, has little to do with the way in which subjects are typically said to fail to meet this condition. Indeed, as we've seen earlier, what McHugh has in mind are rather cases of what he calls “repressed beliefs,” beliefs whose content the subject rejects while maintaining his or her belief. But this is not what is happening in the case of self-deceptive beliefs. It is not that the subject would not react to evidence going against what that subject desires. Successful self-deception in fact requires that the subjects react to what they recognize to be evidence against P , except that they would react contrarily to this evidence. For McHugh, in order to fulfil reasons reactivity, the subject's belief should be such that in a wide range of counterfactual scenarios, the subject would react to evidence “by forming the doxastic attitude she takes them to call for” (McHugh, 2017, p. 2751). In the case of self-deception, evidence recognized as counting against P calls for forming or maintaining the belief that P . Although this way of reacting to evidence is epistemically vicious, it does not show that the subject lacks the capacity to react to evidence and, in this sense, it does not infringe on the subject's responsibility for his or her belief.

The proposition “ S would react to evidence he or she recognizes against P across a wide range of relevant counterfactual scenarios” is true. Therefore, self-deception satisfies reasons reactivity.

VII. CONCLUSION

Self-deceptive beliefs fulfil both reasons receptivity and reasons reactivity. Consequently, self-deceptive beliefs fulfil reasons responsiveness and thus meet a necessary condition for doxastic responsibility. If self-deceptive beliefs are reasons responsive, depending on whether or not one is willing to accept reasons responsiveness as a sufficient as well as necessary condition for doxastic responsibility, then we may be held responsible for our self-deceptive beliefs. However, reasons responsiveness is only a necessary condition for doxastic responsibility. This leaves room for sometimes not being held responsible for self-deceptive beliefs, since reasons responsiveness is not sufficient.

I will not take on the task of supplementing McHugh's account by showing that the reasons-responsiveness condition is also sufficient for responsibility. I do think the latter idea is plausible, however. Before concluding, I would like to at least provide some support for this idea. The following rough picture of responsibility is generally agreed upon: an epistemic condition and a control condition are required and, if fulfilled, add up to responsibility. Each theory then works out these conditions in further detail. We have seen that the case of belief poses a challenge to this picture of responsibility as it is not under our direct voluntary control. Instead McHugh offers a condition for responsibility that depends on an ability the subject should possess. This ability, as we've seen, consists in recognizing alternative evidence to one's belief and reacting accordingly if presented with it. This ability, reasons responsiveness, includes an epistemic component—in requiring a certain awareness on the part of the subject vis-à-vis their reasons for believing—and, for lack of a *control* component, a component that draws on the proper functioning of the subject's mechanism of belief formation. Once we concede that there can responsibility without direct voluntary control, these two components of reasons responsiveness together seem to exhaust what can be required of a responsible belief.

To conclude, I have argued that, at least on Mele's account of self-deception, self-deceptive beliefs always fulfil the reasons-responsiveness condition for doxastic responsibility. This is because the self-deceived subject's relation to evidence is such that the subject would recognize and react to evidence against *P* if that subject were presented with it. I believe not only that this feature of self-deception is present in Mele's account of self-deception, but that we find it in any account of self-deception that posits a similar relation to the evidence on the part of the subject. In this sense, self-deceived subjects always fulfil a necessary condition for doxastic responsibility.

ACKNOWLEDGEMENTS

I am grateful to Anne Meylan, Melanie Sarzano, Benoit Gaultier, and two anonymous referees at *Les Ateliers de l'Éthique/The Ethics Forum* for their constructive and insightful comments, which helped improve this article. I also thank the audiences of the “New Perspectives on Self-Deception” workshop at the University of Montreal, of the “Self-Deception: What It Is and What It Is Worth” conference at the University of Basel, of the “Epistemic and Practical Rationality” workshop at the University of Fribourg, and of the “Ateliers du Groupe de Recherche en Épistémologie” at the Collège de France, and especially Pascal Engel, Veli Mitova, Dana Nelkin, Sarah Stroud, and Christine Tappolet for their helpful questions and comments. Research for this article was supported by the Swiss National Science Foundation (SNSF) Professorship grant “Cognitive Irrationality” # PP00P1_157436 and the Doc.Mobility grant “Irrational, Yet Responsible” # P1BSP1_178606.

NOTES

- ¹ Epistemic at least. I remain agnostic on whether self-deception further consists in a *moral* mistake. In this sense, I follow Neil Levy, who argues that we shouldn't be presumed culpable for self-deception (Levy, 2004). However, I make a step towards disagreeing with him on whether we shouldn't necessarily hold subjects responsible for their self-deception, since I argue that self-deceptive beliefs always fulfil a necessary condition for responsibility. What is being reproached of the subjects is that they should have believed better at least according to epistemic norms. Plausibly, self-deception also constitutes a moral mistake: we might reproach the subjects for believing something they shouldn't according to moral norms as well.
- ² Nelkin rightly points out that, in order to ultimately gain a proper understanding what responsibility for self-deception is, we should “get clear about exactly what the self-deceiver is supposed to be responsible for. We should distinguish between the process of self-deception, the immediate product of self-deception, and its more indirect consequences” (Nelkin, 2013, p. 129). She herself proposes several ways in which one might deal with the questions, one of them being to understand self-deception as a case of culpable negligence. Note also that, as she underlines, her approach to reasons responsiveness is agent based rather than mechanism based (Nelkin, 2013, p. 126).
- ³ Some argue that it is in fact possible to believe at will (Peels, 2015) or intentionally (Steup, 2017).
- ⁴ It is not clear whether McHugh takes reasons responsiveness to be both necessary and sufficient for responsibility. In McHugh, 2013, McHugh mentions that reasons responsiveness might also be sufficient.
- ⁵ In McHugh, 2014, McHugh specifies the kind of reasons in play here. They are what he calls “object-directed reasons”—that is, reasons “for being in some doxastic state that are such because they pertain to whether its content is true or false” (McHugh, 2014, p. 16) or reasons that consist in “considerations that pertain to the world's being as it is represented in the state's content” (ibid., p. 16-17). In other words, when *S* is epistemically reasons responsive, the reasons to which *S* responds are of the sort that make it epistemically warranted for *S* to believe that *P*.

REFERENCES

- Alston, William, "The Deontological Conception of Epistemic Justification," *Philosophical Perspectives*, vol. 2, 1988, p. 257-299.
- DeWeese-Boyd, Ian, "Taking Care: Self-Deception, Culpability and Control," *Teorema*, vol. 23, no. 3, 2007, p. 161-176.
- Fischer, John M., and Mark Ravizza, *Responsibility and Control: A Theory of Moral Responsibility*, Cambridge, Cambridge University Press, 1998.
- Ginet, Carl, "Working with Fischer and Ravizza's Account of Moral Responsibility," *Journal of Ethics*, vol. 10, no. 3, 2005, p. 229-253.
- Hieronymi, Pamela, "Responsibility for Believing," *Synthese*, vol. 161, no. 3, 2008, p. 357-373.
- Levy, Neil, "Self-Deception and Moral Responsibility," *Ratio*, vol. 17, no. 3, 2004, p. 294-311.
- McHugh, Conor, "Epistemic Responsibility and Doxastic Agency," *Philosophical Issues*, vol. 23, no. 1, 2013, p. 132-157.
- , "Exercising Doxastic Freedom," *Philosophy and Phenomenological Research*, vol. 88, no. 1, 2014, p. 1-37.
- , "Attitudinal Control," *Synthese*, vol. 194, no. 8, 2017, p. 2745-2762.
- McKenna, Michael, "Reasons-Responsiveness, Agents and Mechanisms," in *Oxford Studies in Agency and Responsibility*, vol. 1, 2013, p. 151-183.
- Mele, Alfred, "Real Self-Deception," *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 91-102.
- , "Twisted Self-Deception," *Philosophical Psychology*, vol. 12, no. 2, 1999, p. 117-137.
- , *Self-Deception Unmasked*, Princeton, Princeton University Press, 2001.
- , "Self-Deception and Delusions," *European Journal of Analytic Philosophy*, vol. 2, no. 1, 2006, p. 109-124.
- Meylan, Anne, "The Legitimacy of Intellectual Praise and Blame," *Journal of Philosophical Research*, vol. 40, 2015, p. 189-203.
- , "The Consequential Control of Doxastic Responsibility," *Theoria*, vol. 83, no. 1, 2017, p. 4-28.
- Nelkin, Dana, "Responsibility and Self-Deception," *Humana.Mente: Journal of Philosophical Studies*, vol. 5, no. 20, 2012, p. 117-139.
- Peels, Rik, "Does Doxastic Responsibility Entail the Ability To Believe Otherwise?" *Synthese*, vol. 190, no. 17, 2013, p. 3651-3669.
- , "Believing at Will Is Possible," *Australasian Journal of Philosophy*, vol. 93, no. 3, 2015, p. 1018.

———, *Responsible Belief: A Theory in Ethics and Epistemology*, Oxford, Oxford University Press, 2017.

Steup, Matthias, “Belief Control and Intentionality,” *Synthese*, vol. 188, no. 2, 2012, p. 145-163.

———, “Believing Intentionally,” *Synthese*, vol. 194, no. 8, 2017, p. 2673-2694.

Williams, Bernard, “Deciding To Believe,” in *Problems of the Self*, Cambridge, Cambridge University Press, 1973, p. 136-151.