



## Self-deception as omission

Quinn Hiroshi Gibson

To cite this article: Quinn Hiroshi Gibson (2020) Self-deception as omission, *Philosophical Psychology*, 33:5, 657-678, DOI: [10.1080/09515089.2020.1751100](https://doi.org/10.1080/09515089.2020.1751100)

To link to this article: <https://doi.org/10.1080/09515089.2020.1751100>



Published online: 04 May 2020.



Submit your article to this journal [↗](#)



Article views: 313



View related articles [↗](#)



View Crossmark data [↗](#)




Citing articles: 1 View citing articles [↗](#)

ARTICLE



## Self-deception as omission

Quinn Hiroshi Gibson 

Department of Philosophy, American University of Beirut

### ABSTRACT

In this paper, I argue against three leading accounts of self-deception and propose a heretofore overlooked route to self-deception. The central problem with extant accounts is that they are unable to balance two crucial desiderata: (a) to make the dynamics of self-deception (e.g., the formation of self-deceptive beliefs) psychologically plausible, and (b) to capture self-deception as an intentional phenomenon for which the self-deceiver is responsible. I argue that the three leading views all fail on one or both counts. However, I claim that many or most cases of self-deception conform to a different model, which I call ‘self-deception as omission.’ In these cases, the process of self-deceptive belief formation and the intentional act for which the self-deceiver is responsible come apart, allowing us to meet both desiderata. Self-deceptive beliefs are often formed by unconscious mechanisms closely analogous to “System 1” processes of dual-systems psychology, or by other mechanisms of motivated reasoning. The nascently self-deceptive subject then acquiesces in the comforting belief and commits an epistemic failure by allowing it to persist. If this is done for motivationally biased reasons – for example, preferring that the belief in question be true – then the subject is self-deceived and is blameworthy for her epistemic omission.

### ARTICLE HISTORY

Received 10 January 2018  
Accepted 2 July 2019

### KEYWORDS

Moral responsibility; dual-process theory; self-deception; moral psychology; irrationality; motivation; motivated reasoning

## 1. Introduction

It is very natural to think of self-deception as a special variety of ordinary deception, where the deceiver and the deceived just happen to be identical. In ordinary deception, the deceiver intends to get the deceived to believe something that the deceiver believes to be false. We might call the view that results from thinking of self-deception a special case of ordinary deception, the “naïve” view of self-deception:

**The naïve view of self-deception:** A is self-deceived that p just in case A believes that not-p, and A has acted intentionally so as to cause A to believe p.

However, on inspection, this starts to look like something that no one could ever manage to do successfully. Difficulties with having contradictory beliefs

aside,<sup>1</sup> how is an agent to act intentionally so as to get herself to have a belief when she also already believes the object of that very belief to be false? The problem is not merely that belief is not under the control of the will. Indeed, there may be nontrivial ways in which belief-formation is under the control of the will. However, I take it that no defensible version of doxastic voluntarism would allow that one can believe just anything at all at will. In addition, if anything provides constraints on what one can believe, it ought to be the other beliefs one has, evidence for which has already been appreciated.

The biggest problem, then, with the naïve view of self-deception seems to be that it describes an act that – according to doxastic non-voluntarists and (moderate) voluntarists alike – isn't possible. Intending to form the belief that *p*, and succeeding, seems to require that one have evidence for *p* (according to non-voluntarists), or at least, that one does not have evidence for not-*p* (according to moderate voluntarists). However, both of these conditions are violated by a self-deceiver if the naïve account of self-deception is correct, given that she already believes not-*p*. The trouble with the naïve view does not spring from the idea of belief, nor from the idea of intentional action per se, but rather, with the particular way in which they are said to interact in the mind of a single more-or-less unified agent. There are three moving parts to the problem: belief, intentional action, and psychological unity. If the self-deceived agent could somehow pull off an intentional act of getting herself to believe what she also believes to be false, wouldn't this undermine her psychological unity? If what the self-deceived, sufficiently-unified agent manages to bring about in herself is a genuine belief, wouldn't she have to have done it "non-intentionally"? *Mutatis mutandis*, if the sufficiently-unified self-deceived agent really intends to deceive herself, how can the deception involve the bringing about of a genuine belief? Let us call this difficulty for the naïve view of self-deception, following A. Mele (1997), the "dynamic problem" of self-deception.<sup>2</sup> With this problem in focus, the central philosophical question about self-deception becomes this: how is it psychologically possible? Much of the philosophical literature on self-deception is organized around trying to find a solution to the dynamic problem, and the available views can be sorted according to how they depart from the naïve view. Understanding the problem in this way helps us to see the three main strategies that various philosophers have embraced to solve the dynamic problem:

- (1) Deny that one of the beliefs in the contradictory pair rises to the status of full belief
- (2) Deny that the "self" involved in self-deception is unified; and
- (3) Deny that self-deception is fully intentional

I choose to deal with the dynamic problem in a different way. Rather than trying to tweak the content of the self-deceptive intention or the precise nature of the representational state that the self-deceiver is in, I propose that the structure of the self-deceptive process – at least in many cases – needs to be reanalyzed. Oddly, many theorists who have tried to take seriously the dynamic problem that the naïve account seems to face have not thought to alter the form of agency that the naïve account imputes to self-deceivers – that is, roughly, intentional action of some kind. According to my view, self-deception need not involve intentional action, but rather, it more often involves intentional failure to act, intentional omission. Acknowledging a path to self-deception through this kind of failure solves many of the outstanding problems with other views, and though I do not claim that to be the only path to self-deception, it has thus far been overlooked.<sup>3</sup> My account is constrained by – and motivated to capture – another very important aspect of self-deception. Self-deception is more than just willful belief-formation. It is also a type of self-induced ignorance, and there is, thus, a very strong *prima facie* presumption that self-deceivers are responsible for their self-deception. This is, in part, what makes self-deception a phenomenon of concern for moral psychology: can we make good sense of self-deception in a way that justifies the attitudes that we hold toward those who perpetrate it on themselves? Ordinary moral thinking seems to hold that self-deceivers are responsible and that blame is *prima facie* appropriate. If there is one important thing that the naïve view gets right, it is this: there is a recognizable form of agency, namely, an intentional action on which we can hang our judgments of blameworthiness. Solving the dynamic problem thus ought to go hand in hand with preserving the idea that self-deceivers are responsible. However, as I will try to show as we go along, the competing goals of responding to the dynamic problem and holding our responsibility judgments intact are in tension. The way to resolve the tension, and to satisfy both desiderata, is to find the locus of agency in self-deception – not in an intentional action which is identical with the act of deceiving oneself, but instead, in an intentional omission, which only partially constitutes the process of self-deception as a whole. First, I will outline my view in more detail and show how it both avoids the dynamic problem and holds intact our responsibility judgments. Then, I will further elaborate the advantages of the view with relation to other views in the literature. I will argue that my view escapes objections – both new and old – that plague other views.

## 2. Self-deception as omission

According to my account, which I call self-deception as omission, the episode of intentional agency for which the self-deceiver is responsible and the process of belief-formation come apart. According to this view, the agent is not responsible for the formation of the self-deceptive belief itself, but rather,

they are responsible for acquiescing in that belief once it has been formed, where the acquiescence itself is a voluntary omission. Therefore, self-deception, following my view, is a two-part phenomenon. First, there is the process of belief formation. I will elaborate on this stage extensively below, appealing to the empirical literatures on biases and dual-processing, on the one hand, and the mechanisms of motivated reasoning, on the other hand.

The second stage is the maintenance of an ill-supported or defeated belief via a motivated failure to investigate the matter further. Self-deceptive belief concerns some matter about which the agent has a desire. According to my view, if the desire that a certain defeated belief be true, and causes the agent to persist in the belief by motivating forbearance from further investigation, the agent is self-deceived. That is, if evidence against the belief is available – in the sense that a reasonably judicious look into the matter would reveal it – but the agent is caused to forebear from further investigation by a desire to remain in an affectively more palatable state when she otherwise would proceed, we have a case of self-deception.<sup>4</sup> This is how we solve the dynamic problem: the belief can come about in the agent even though she never intends to get the belief to come about in her, and we are able to get this result without doing violence to the intuition that self-deception is something that the agent had perpetrated on herself. What is irrational about the agent accepting the belief is that she acts as though the cause of her belief and its effects provide good reasons to continue to hold it. The thing for which she is responsible is her acquiescence in the affectively more palatable but epistemically unwarranted state.

Therefore, according to this view, there is no single act which is an act of intentional belief-formation against a belief that one already holds. One may, of course, hold the negation of the self-deceived belief, but I don't think there is anything particularly puzzling about being in that state. What is puzzling, rather, is how one can manage to get into it intentionally. The answer is that one doesn't get into it intentionally, but remains in it intentionally. Decomposing self-deception into this complex two-part phenomenon solves the dynamic problem without failing to capture the way in which self-deceivers are responsible for being self-deceived. We can state my view as follows:

**Self-deception as omission:** An agent is self-deceived that *p* if she believes *p* and intentionally omits to seek, recognize, or appreciate externally available evidence for not-*p*, for reasons which ultimately derive from her desire that *p* be true, in a way which enables the maintenance of the belief that *p*.<sup>5</sup>

This principle only states a sufficient condition for self-deception; I do not claim that there are no other ways that one can be self-deceived. In fact, I think there are good reasons to think there are other ways to be self-deceived (see the discussion of Mele, below). However, I do claim that this is a common route to self-deception that has been overlooked. As I hope will

become clear, the range of other possible routes to self-deception is significantly constrained by the dynamic problem – constrained much more than what is acknowledged by other theorists – so uncovering a sufficient condition that is not subject to that problem is of theoretical interest.

Strictly speaking, my view as stated does not require that the self-deceptive belief be formed in any particular way at all, but, because my view embeds an epistemic requirement – that there be available evidence for not-*p* – belief-forming processes that are more likely to lead to forming beliefs that are defeated in this way will be especially likely to set agents up for self-deception. These belief-forming processes also interface importantly with large scale cognitive architecture that is relevant for assessing whether behavior is intentional. Allow me to elaborate.

We now know that many belief-forming mechanisms are biased. For example, consider the availability heuristic. When tasked with making judgments of probability or frequency, people are often led seriously astray by giving undue weight to things that are easier to recall or are more salient. While it's true that if an event is more frequent, it will be easier to recall, the converse does not hold. Nevertheless, the availability heuristic operates in accordance with such an invalid principle. In a famous experiment, Tversky and Kahneman (1973) asked subjects to judge the relative frequency of words beginning with the letter 'k' against the frequency of words with 'k' in the third position. Subjects consistently judge that words beginning with 'k' are more frequent, despite the fact that they occur only somewhere between one third and one half as often as words with 'k' in the third position. Tversky and Kahneman concluded that this was because when faced with the task, subjects set about recalling words in each of the two categories. Since it is much easier to recall words that begin with the letter 'k,' subjects concluded that those words were more frequent.

Psychologists now recognize countless biases that have a similar structure, and a highly general theory of cognitive architecture has emerged which is designed to account for them. So-called “dual-system” or “dual-process” theories of cognition – in addition to being elegant, explanatorily powerful, and robustly empirically supported – can help us understand the structure of self-deception. Dual-process theories of cognition posit a basic division of labor within the mind between two subsystems which accounts for the experimentally observed biases.<sup>6</sup> The first, System 1 (S1), is fast, intuitive, nondeliberate, subconscious, and can work in parallel. The second, System 2 (S2), is slow, effortful, conscious, and serial. Consider the following example by Kahneman (2011) to illustrate S1 at work. His instructions are to “not try to solve it, but listen to your intuition” (p. 44):

A bat and a ball cost \$1.10. The bat costs one dollar more than the ball. How much does the ball cost?

The intuitive – incorrect – answer that System 1 offers up is \$0.10, and it seems to do so more or less unbidden, once the specification of task has been grasped. Whether one chooses to go on and perform the calculation and ultimately arrive at the correct answer seems to be an independent matter. What happens is that, first, S1 issues its verdict, and then – if one chooses – one can go on to perform the calculation and then override the initial judgment offered up by S1. The initiation of S2 is voluntary, and its operations are effortful, but sometimes they are necessary to avoid error.

At least some of the time, S1 aims at the production of beliefs,<sup>7</sup> but what the results from the empirical literature appear to show is that S1 aims not just at the production of true beliefs, but at the rough-and-ready production of for-the-most-part true beliefs. It is widely believed that the evolutionary value of a fast and highly adaptive parallel system outweighs the disvalue of sometimes getting it wrong when it comes to, for example, abstract problems about probability. When the goal is to believe in accordance with the norms of probability theory, one must critically scrutinize the immediate offerings of intuition. What leads the subject astray is an insufficiently judicious look at a belief that she has been automatically saddled with.

This cognitive architecture interfaces in interesting ways with known processes of motivated reasoning. Psychologists usually think of motivated reasoning as evidential search and evaluation biased by goals: “reasoning involves the recruitment and evaluation of evidence. Goals can distort both of these basic cognitive processes” (Epley & Gilovich, 2016, p. 136). In general, theories of motivated reasoning can be divided into “quality-of-processing” accounts and “quantity-of-processing” accounts. According to quality-of-processing accounts (e.g., Kunda, 1990), goals bias reasoning by affecting the range of hypotheses selected for testing and the inferential rules that are used in the evaluation of those hypotheses. According to quantity-of-processing models (e.g., Ditto & Lopez, 1992), agents are simply more motivated to process preference-inconsistent information: “wants and fears may often bias judgments ... because of the simple fact that preference-consistent information is accepted ‘at face value,’ whereas preference-inconsistent information tends to trigger more extensive cognitive analysis” (Ditto & Lopez, 1992, p. 581).

I will not attempt to settle which of these accounts is correct. Indeed, they may be complementary. That is, motivated reasoning may involve both kinds of mechanisms in different cases, and either kind of motivated reasoning could be playing a role in self-deception.

First, let us consider the quality-of-processing accounts. Sometimes, the qualitatively different processing that is initiated under the influence of a goal is just S2 processing. For example, in one study subjects were given a standard scenario where base rates are typically ignored, such as

Kahneman and Tversky's (1972) "cab problem." In this problem, subjects are told that a witness reports that a green cab was involved in a hit-and-run. They were also told about the reliability of the witness and the prior probability that the cab would be green, and were asked to estimate the likelihood that the cab in the accident was green. However, when asked to answer as though they were a lawyer for the green cab company, subjects used the base rate information much more reliably and produced lower estimates when the base rate was low (Ginossar & Trope, 1987). In this study, the manipulation gives the subject a specific accuracy goal which reduces the effect of ignoring base rates (a bias in S1 processing) by prompting subjects to consciously take account of base rates, namely, to engage in explicit S2 processing.

In other cases, the qualitatively different kind of processing that is induced under the goal is simply biased S1 processing. For example, there is considerable evidence that subjects can be induced to perform a biased memory search for evidence that they possess a specific trait (e.g., introversion vs. extraversion) if they have been made to believe that that trait is desirable (e.g., that it correlates with business success) (Santioso et al., 1990).

One observation that has been made about quality-of-processing views is that they seem to require that the agent, "at some level, already 'know' what an inference strategy will yield before choosing to use it" (Ditto & Lopez, 1992, p. 581). This might seem to suggest that we have most of the materials present for an account of self-deception already in hand. However, this understanding of the view seems to rest on the assumption that the initiation of qualitatively different processing is done intentionally by the agent. However, regardless of whether the induced processing is of the S1 or S2 variety, the assumption could be false. If it is the former, the entire process – from selection of the mechanism through to the actual processing – can take place in S1. If it is the latter, the selection of the relevant mechanism is not done knowingly by the agent, even if, in some sense, the relevant processing is. That is to say, motivated reasoning construed according to a quality-of-processing account still falls short of self-deception. What self-deception requires is more direct involvement from the agent and the agent's motivations over a period of time; my view cashes this out with the requirement that the agent and her motives can properly be said to be responsible for the maintenance of an ill-supported belief. What goal-induced processing (either of the S1 or S2 variety) does is set the subject up for self-deception by issuing a belief which may or may not be on a secure foundation. If the belief is, in fact, false or poorly justified, the subject becomes self-deceived by failing to notice that the belief has this status if the reason for forswearing further investigation is the subject's preference for continuing to believe as she does.

If I am right, then in many cases, the self-deceptive belief will be formed as the result of the operation of biased processing; but my view doesn't require that it be formed in that way. The mechanisms of motivated reasoning are good candidates for playing a crucial role in this stage of the process because they will often be biased, resulting in ill-supported beliefs. However, my view departs from the naïve view precisely by finding what is distinctive about self-deception, not in the process of belief-formation, but in the dynamics of belief maintenance. It is often taken for granted in the self-deception literature that nothing should count as a self-deceptive state that does not arise from some distinctively self-deceptive process. Although my view denies that this connection is necessary, it is important to note that many of our mental processes (such as those highlighted by dual-process theory) are of a sort that makes us vulnerable to forms of epistemic failures like self-deception. Biased processes can lay the groundwork for a later episode of mental agency to bring about a genuinely self-deceptive state. Goal-induced processing is a particularly good candidate for producing nascently self-deceptive beliefs because the desire that motivates the subsequent epistemic omission comes neatly packed into the process of belief formation itself.

Let us now consider the quantity-of-processing account. According to this account, "the central way that motivational factors affect judgments is through their effects on how extensively preferred and nonpreferred information is analyzed" (Ditto & Lopez, 1992, p. 580). As Epley and Gilovich (2016) put it:

People . . . ask themselves very different questions when evaluating propositions they favor versus oppose . . . When considering propositions they would prefer to be true, people tend to ask themselves something like "can I believe this?" . . . When considering propositions they would prefer not to be true, people tend to ask themselves something like "Must I believe this?" (p. 137)

That is, the evidential standard that must be met is higher when entertaining nonpreferred conclusions than it is when entertaining preferred conclusions. The agent is motivated to do more effortful processing when the evidence is potentially threatening, and comfortable conclusions are much more easily accepted.

A typical experiment in the paradigm goes like this. Participants are told that they are going to be given a test for the presence of some factor (a deficiency in a made-up enzyme) that correlates some with serious disease. They are then "tested" for the deficiency. Every participant observes the same result (no change in a litmus strip in response to saliva), but one group is told that this result indicates the deficiency, while the other is told it indicates normalcy. Participants who were led to believe that they had the deficiency kept the testing paper in the cup containing saliva significantly longer, reinserted the litmus strip into the cup, put it directly in their mouths, and

tried “shaking, wiping, blowing on, and, in general, quite carefully scrutinizing the recalcitrant . . . test strip” (Ditto & Lopez, 1992, p. 576).

One might sensibly wonder whether, according to my view, the subjects who easily accept the favorable results are guilty of self-deception. They are guilty of a motivated failure to gather further evidence; all of the extra processing that the “deficient” subjects engaged in is clearly of the S2 variety, and we may even suppose that the subjects who engaged in it are responsible for initiating it. We might then conclude that the subjects who failed to engage in it are responsible for that failure in a way that suggests they satisfy my account.

My view, however, does not classify these subjects as self-deceived. The reason is simple: the situation they are in is not one where further investigation is warranted. The belief that they have (i.e., that they are healthy) is both true and justified. There is no further available evidence that their motivated reasoning causes them to overlook something because there is no further available evidence. Moreover, the evidence that the subjects take themselves to have is of a variety that is normally reliable. Now, of course, the “evidence” that they have bears no real connection with the truth of what they believe, though what they believe is true. However, they are not guilty of the kind of motivated epistemic omission that self-deception requires. This feature of my view is not merely stipulative. One shouldn’t be epistemically required to do more than what is required for justification. Therefore, if the belief is justified, even luckily, one can’t be guilty of a blameworthy epistemic omission. Since my view requires such an omission for self-deception, it has the consequence that one cannot be Gettiered and self-deceived at the same time.

There will, however, be some cases where agents engage in motivated reasoning, construed according to a quantity model, where they are truly self-deceived. The quantity account, unlike the quality account, does not terminate in the formation of a belief. According to the latter accounts, the episode of motivated reasoning is over when the agent has settled the matter in question. For this reason, we need the second stage if the agent is to be truly self-deceived. However, quantity accounts take us much further. Suppose a subject who actually had a disease were undergoing a test for that disease which indicated a positive result by the litmus paper turning color. Suppose, further, that the test often yields a false negative on the first trial, but that repeat trials are reliable. Suppose the subject gets a false negative, and on this basis, forms the belief that she does not have the disease. At this point, she is not yet self-deceived, she has merely been misinformed by an unreliable test. However, if she then fails to seek further evidence about the reliability of the test for motivationally biased reasons (as the quantity account says she may very well), she is self-deceived.

What is crucial is that no matter how the motivated reasoning plays out exactly, there must be a substantial motivated contribution by the agent at

the personal level for her to count as self-deceived. Belief-formation itself is not sufficiently agential to satisfy this requirement. This is the essence of the dynamic problem, and competing views do not fully avoid it. Finding the episode of agency downstream of belief formation is the most promising way to avoid this problem (see [Section 3](#)).

The importance of the contribution of motivation needs to be emphasized. Plain old bad reasoning is not sufficient for self-deception, according to my view, nor is being subject to one of the myriad cognitive biases that psychologists have uncovered. The operation of those biases makes subjects vulnerable to self-deception in the way that a genetic predisposition might make someone vulnerable to developing a disease; having the predisposition is no guarantee one will develop the disease, and there is much that one can do to stave it off. Unfortunately for us, some of our epistemic vulnerabilities are quite widespread, so a great deal of vigilance may be required to avoid lapsing into self-deception, which is, as experience attests, all too easy. It is therefore fitting, I think, that self-deception as omission makes self-deception quite common; I think it is quite common. A detective who makes up her mind early in the investigation that a certain party is guilty might not be self-deceived if she is merely manifesting confirmation bias, but she will be guilty of self-deception if there is evidence against her hypothesis that is available, and a motivation – say, a motivation to see that party convicted, or even a motivation to have the investigation closed – is preventing her from seeing it. A gambler reasoning with her namesake fallacy isn't self-deceived unless some motivation of hers is playing a similar role. There will be many cases like these, and I consider it a feature of my view, rather than a bug, that it counts them as cases of self-deception.<sup>8</sup>

Since I have been emphasizing that it is important for any account of self-deception to get the facts about responsibility right, I should highlight that, according to my view, that for which the self-deceived agent is responsible is a voluntary omission. Now, not all omissions are straightforward things for which an agent is fully responsible. In typical cases where robust responsibility attaches to an omission, the agent is aware of the consequences of failing to act, as well as the alternatives – actions – and their consequences (within reason), yet chooses to refrain from performing any of them. Suppose I am minding my own business on a park bench as a trolley whizzes past. Suppose, further, that at that very moment a strong breeze parts the trees in front of me, revealing both a helpless bystander trapped on the track and a lever with a sign reading “Flip Switch to Divert Trolley.” It seems to me that I would be at least partially responsible for the bystander's death if I didn't at least try to get to the lever in time. Contrast this with the case where I am aware of neither the bystander nor the lever. Assuming my ignorance is itself excusable, it is that ignorance which relieves me of responsibility for failing to act to save the bystander.

Thus, one might worry whether any of the conditions that typically attend cases of robust responsibility for omission are present when an agent is self-deceived, according to my view. It is precisely in response to the dynamic problem attending the naïve view that we are motivated to find a view where the agent isn't aware of what she is doing as an act of deceiving herself. Doesn't this undermine her responsibility?

I grant that in typical cases of omission, certain awareness conditions must be satisfied by the agent. The agent cannot be held responsible for failing to act on an alternative she was excusably ignorant of. This seems to follow from some version of "ought-implies-can": the agent's knowledge of the nature and availability of the alternatives is a condition on her choosing to act on any of them; but the requirement that the self-deceived agent violates is not one that fails to apply in cases of ignorance. The self-deceived agent is not excusably ignorant of what she is ignorant of – she has acquiesced in a comfortable but evidentially defeated belief. Exercising epistemic agency in accordance with the norms that govern it is an effortful process, and the results can be disappointing, or worse – but this does not excuse the agent from doing it. Thus, a failure to do so – motivated by a desire to avoid effort, to avoid possible disappointment (or worse), and to remain in the comfortable ruddy glow of one's positive self-image (as the case may be) – is a failure for which the agent is blameworthy. Ignorance may be enough to change how one is required to act in a given situation, but it cannot itself excuse the agent from the epistemic norms which determine whether that ignorance is itself blameworthy or not.<sup>9</sup>

My view of responsibility for self-deception rests on a form of epistemic deontology. That is, what the self-deceiver is guilty of is failing to reason as she ought to from evidence which is in her power to collect and appreciate. My response to the worry that responsibility for omissions requires the satisfaction of awareness conditions that are not satisfied in cases of self-deception is, thus, that these conditions do not hold of the epistemic norms that the self-deceiver violates.

This response, however, raises another worry. Does it turn out, according to this view, then, that the self-deceiver is not morally blameworthy, but merely epistemically blameworthy? My response to this is that there is no hard and fast distinction between moral norms and epistemic norms. Indeed, epistemic deontologists may help themselves to explanations for the normative force of the principles of epistemic conduct which appeal, in part, to the moral significance of doing and not doing one's epistemic duty. For example, explaining why he thinks "it is wrong always, everywhere, and for anyone to believe anything on insufficient evidence," Clifford (1999) says:

If I let myself believe anything on insufficient evidence, there may be no great harm done by the mere belief; it may be true after all, or I may never have occasion to exhibit it in

outward acts. But I cannot help doing this great wrong towards Man, that I make myself credulous. The danger to society is not merely that it should believe wrong things, though that is great enough; but that it should become credulous, and lose the habit of testing things and inquiring into them; for then it must sink back into savagery. (p. 76)

What I wish to borrow from Clifford is the justification of the principles of epistemic deontology by appeal to genuinely moral considerations in order to secure the moral wrongness (in at least some cases) of failing to do one's epistemic duty. Although a persuasive defense of this position would take us significantly beyond the scope of the current discussion, I am attracted to a pluralistic appeal to moral considerations for the justification of the principles of right epistemic conduct. In the passage just quoted, while admitting that insufficiently supported beliefs can have bad consequences, Clifford seems to also be appealing in part to a virtue-ethical explanation for the moral wrongness of believing on insufficient evidence, that it leads to the cultivation of credulity. The reason why credulity is, in turn, morally problematic is explained in terms of its social effects. Whether credulity is also supposed to be problematic in itself, or whether the social effects of the cultivation of credulity are ultimately bearers of disvalue, Clifford seems to leave open. I make no claim about what Clifford's considered position is, but it seems very plausible that we have a great variety of morally significant reasons to be concerned with agents' epistemic conduct. Sometimes, failing in one's epistemic duty increases the risk of harm to the self and others, and perhaps it leads to the cultivation of inherently problematic epistemic vices.<sup>10</sup>

Thus, my response to someone who is worried that self-deceivers are responsible only for epistemic misconduct and are not morally responsible is that some of the principles of epistemic conduct (whatever they are precisely) have genuinely moral justifications. That is a controversial position, but my view of self-deception as such does not rest on it. What rests on it is whether the self-deceiver is morally blameworthy, in addition to being epistemically blameworthy. I am inclined to think that self-deceivers are morally blameworthy, but for those who are not inclined to agree with me on this, self-deception as omission still has something to offer: whatever kind of blameworthiness we think attaches to self-deception, we need a view that gets the result that self-deceivers are responsible without making it psychologically perplexing, and self-deception as omission can deliver that. The kind of responsibility that results will depend on how we conceive of the relation between the epistemic requirements that the self-deceiver violates and moral requirements. These concerns are related to an argument that has been given by Neil Levy. Levy (2004) questions whether "new" accounts (according to which self-deception is not intentional) can capture, or should even try to capture, the responsibility of self-deceivers. He claims that, while traditional accounts (he has in mind type (2) accounts, above, as well as others) imply that self-deceivers are responsible,<sup>11</sup> if we abandon accounts of that kind, we

should also abandon the “presumption” that self-deceivers are responsible. His argument is this: responsibility judgments require both that the subject matter at hand be morally significant, and that the subject have some degree of counterfactual control over the relevant conduct. Applied to cases of epistemic failure, these requirements become: (1) that the matter of believing or not believing the thing in question is morally significant, and (2) that “we are in some doubt about” (p. 305) the truth of the belief in question. However, Levy claims that these are conditions that are not likely to be satisfied by very many cases of self-deception because “successful acts of self-deception [according to a non-intentional model] leave me in no doubt about the proposition concerning which I am self-deceived” (p. 307). Levy seems to be assuming that being in a state of doubt about the truth of one’s belief is the only way one can begin to exercise control – the kind of control relevant for moral responsibility – over whether one holds that belief or not. If doubt is understood to be an occurrent state of ‘wondering whether p,’ or even as ‘failing to treat the matter of whether p as settled,’ this is surely false. If an S1 process offers up the wrong answer to the bat-and-ball problem – or the Linda problem; Kahneman and Tversky’s questions about availability; or, more controversially, activates a racist or sexist schema at the sight of someone – there may be no occurrent doubt, in the relevant sense, whatsoever about whether my answer or judgment is correct until S2 has been activated. Nevertheless, it is in my control whether those beliefs persist in me and become springs of action because it is in my control whether to activate System 2 or not.<sup>12</sup> If this is correct, then doubt is not required for responsibility, even though it may be true that a doubt-entailing alternative course of action must be available. On the other hand, if doubt is simply understood to be whatever state it is that precedes and enables changing one’s mind about p, then there is no reason to think that self-deception – certainly as understood according to self-deception as omission – should rule out that the subject can be in it. Ought-implies-can is not violated for cases like these just because the subject has a more-or-less settled view of the matter.<sup>13</sup>

### 3. Comparison and elaboration

According to self-deception as omission, it is not because of anything that the agent does, but rather, because of something that the agent fails to do, that the agent ends up with, and is responsible for having, a self-deceptive belief. This feature of the account distinguishes it from any other that I know of.

As I mentioned above, views of self-deception can be sorted according to how they depart from the naïve view. Recall that the three prominent strategies are these:

- (1) to deny that one of the beliefs in the contradictory pair rises to the status of full belief;
- (2) to deny that the “self” involved in self-deception is fully unified; and
- (3) to deny self-deception is fully intentional.

This categorization is not meant to be exclusive, however. Views of self-deception may cross the boundaries of these categories. Indeed, self-deception as omission is a view of both type (2) and type (3). I would like to further elaborate my view, distinguishing it from other views of these kinds.

### 3.1. *Mele*

Self-deception as omission is of type (3). It denies that self-deception, as a whole, is fully intentional because a crucial part of it – the stage where the belief is formed – is not intentional at all. Al Mele’s influential account is also of this type, but there are crucial differences between Mele’s view and my own. Mele’s strategy is to solve the dynamic problem by claiming, in effect, that no one deceives themselves *de dicto*, but only *de re*. According to Mele (1997, 2001), in cases where I deceive myself, I cause myself to believe something false by treating the evidence in a motivationally biased way. I never intend to deceive myself, merely to redirect myself away from certain kinds of evidence. How does this view compare to self-deception as omission? Does it solve the dynamic problem? Is it psychologically plausible?

One of Mele’s examples is that of Beth, a 12-year-old girl whose father has recently died. She has come to form the belief that her father loved her most of all his children. She has come to this belief by selectively attending to pleasant memories of her father playing with her alone, and selectively ignoring those memories of her father playing with her brothers. Her evidential selectivity is explained by a motivation – a motivation to attend to pleasant memories over unpleasant ones – but that motivation is not a motivation to deceive herself.

This case is compelling on its face. Beth’s case does appear to be one where she acts in a non-paradoxical way that leads her to a self-deceptive belief. The question to ask at this point is whether the general dynamic problem will not arise again, for at least some cases. The case of Beth has some plausibility because the evidence that she is led by bias to entertain (i.e., the comforting memories of her father playing with her) has some psychologically pleasant quality to it, which is independent of the conclusion that it warrants; similarly, the evidence that she avoids (i.e., the times her father doted favorable attention on her brothers) has an unpleasant quality independent of the conclusion that it warrants. However, is it plausible that all cases can be assimilated to this model?

This worry for Mele's view has been persuasively articulated by Robert Lockie (2003). If someone only finds evidence (un)pleasant in relation to the (un)pleasantness of what it is evidence for, then her avoidance of the unpleasant bunch is either motivated and purposive, or it is not. For it to be a genuine case of self-deception, it seems it must be motivated and purposive. However, we then seem to have a version of the dynamic problem again, for how can she be motivated to avoid evidence if the aversive character of the evidence derives from the aversive character of the thing which it is thought to be evidence for – the very thing about which, if she purposively avoids, she is guilty of purposively deceiving herself?

Mele's view still requires self-deceivers to act in such a way that is guided by the aversive quality of the evidence. There are some cases, like the Beth case, where this is psychologically unproblematic because the evidence is intrinsically aversive. However, many more typical cases of self-deception are unlike the Beth case; they involve evidence of the aversive character which derives from an appreciation of evidence – and what it is evidence for – in a way that is not consistent with plausible psychological dynamics.

I agree that Lockie's worry is a problem for Mele, but this problem does not arise with self-deception as omission. According to my account, there is no need for an agent to be motivated to avoid particular evidence. Motivation is still playing a role according to my account, but it comes in at a later stage, playing a role in maintaining the belief, rather than in the formation of the belief itself – so there is no need to think that the agent has already interpreted the evidence against the belief as threatening. According to self-deception as omission, there is no action the explanation of which requires us to think some evidence has an aversive quality, either intrinsically or derivatively. What has an affective quality (albeit a positive one) is the belief with which the agent already finds herself, and what she does – or rather, omits to do – is explained by that belief and its attendant affect itself. There is nothing mysterious about why the belief in question has the attendant quality that it does: it is just more pleasant to believe in accordance with what I desire than to fail to so believe. This contrast – between what is more and less pleasant to believe – does not have to be entertained by the agent, according to my view. All we need to find is a reason why the agent continues to believe as she already does – that is, we only need to find a reason why the agent fails to investigate further. This does not require a contrastive second-order belief representing alternative beliefs as more unpleasant to hold; it merely requires that the agent be carried along by the pleasant inertia of her current state.

My view does undoubtedly share some features with Mele's view. We both agree that the dynamic problem ought to be taken seriously, and we both deflate the intentional nature of self-deception (according to the naïve view) in order to avoid it. However, Mele still identifies an act of self-deception, and

the psychology to which he appeals to make that act plausible, in some cases, does not generalize (à la Lockie). I therefore do not think that Mele's view offers a general solution to the dynamic problem. That is, unless Mele is willing to significantly restrict the range of cases of bona fide self-deception, he has not identified the necessary conditions for self-deception. Self-deception as omission doesn't identify the necessary conditions for self-deception either, but it does identify a range of possible routes to self-deception that would be closed if Mele's view were correct, and it does this by describing a process that is free from the dynamic problem.

### 3.2. *Fracturing*

Lockie's objection to Mele's view is given in service of his argument that no account of self-deception which isn't "depth-psychological" will adequately capture the phenomenon. For Lockie (2003), a view is "depth-psychological" if it involves "a person of parts, these parts with their own motivational interests, whose activities and motivations are not necessarily known to, or shared by the other parts" (p. 137). According to the classification of views that I have given, a depth-psychological view is of type (2), that is, is of the type which adopts the idea that the self is in some important sense less than perfectly unified. Self-deception as omission is also a view of this type. I do, after all, rely on the idea that some cognitive operations go on independently of one other, and certainly without the conscious awareness of the agent. However, self-deception as omission is not a depth-psychological view, in Lockie's sense of that term. I do not have anything against depth-psychology as such, but I do not think that we need it to have an adequate account of self-deception.

Views of this type have always been popular explanations for the dynamics of self-deception.<sup>14</sup> If we locate both the belief that *p* and the motivation to believe not-*p* in the unconscious, we seem to be able to explain how the agent can have seemingly contradictory beliefs (if we are worried about that), and to offer a solution to the dynamic problem of self-deception at a single stroke. Perhaps one advantage to this kind of approach is that it leaves intact our ordinary notions of belief and intention, and instead, puts pressure on the idea of a unified self, which we might have independent reason to reject anyway. By maintaining that self-deception is intentional, fracturing views may also appear to be well-positioned to capture the responsibility of self-deceivers. This, however, turns out not to be so straightforward.

One major division between fracturing views is whether the parts of the agent are themselves thought of as loci of agency. According to a view such as Davidson's, the only division required is a form of inferential isolation – or what Davidson calls a "boundary" – between inconsistent beliefs, whereas according to Lockie's view, the parts of an agent are much more robust.

However, even according to Davidson's view, agency enters the picture because the agent as a whole is playing a role erecting and maintaining the boundary. He says that "the irrational step is therefore the step that makes [self-deception] possible, the drawing of the boundary that keeps the inconsistent beliefs apart" (Davidson, 2004a, p. 211). On the assumption that self-deceivers are responsible for that which is irrational about self-deception, it seems that Davidson's view can only capture the responsibility of the self-deceiver if the agent herself is responsible for erecting and maintaining the boundary. However, how can it be the agent herself who draws the boundary without running afoul of the dynamic problem after all? Is the drawing of the boundary motivated and purposive or not? Why is it drawn there?

While a view like Davidson's seems unable to capture the responsibility of self-deceivers without falling into the dynamic problem, depth-psychological views face more direct problems having to do with responsibility. These worries are well-documented (e.g., Johnston, 1988). Levy (2004) describes the worry this way:

The [depth-psychological] account splits the mind into a deceiving system and a deceived system. It therefore absolves the deceived system of all responsibility for the deception. On most such accounts, the deceived system is identified with consciousness; with the agent's real self. Thus, on [such an] account, the agent is not responsible for self-deception. At this point, *modus tollens* kicks in; since we know that self-deceivers are responsible for their self-deception, [depth-psychological] accounts are false. (p. 302)

Levy doesn't think this argument will work because he thinks the "presumption" that self-deceivers are generally responsible is false. However, since I think that self-deceivers are generally responsible, and since I have already offered my reply to Levy's argument against this, above, we can take this argument as a significant strike against depth-psychological views. However, as with Lockie's objection to Mele, self-deception as omission escapes this problem. The only agent in the picture, according to my view, is the person herself. There is no subagent which perpetrates the self-deception, according to my view. The mind has parts, if one likes to put it that way, but those parts are not loci of agency. It is only the person herself who fails, for motivationally biased reasons, to do what is required to bring belief into conformity with epistemic requirements.

### 3.3. *Pretense*

I have just distinguished self-deception as omission from other views of type (2) and type (3) (to which self-deception as omission itself belongs) by arguing that it is not vulnerable to objections that have been previously leveled against views of those types. Now, I would like to turn to the remaining family of views, one

variety of which has recently been gaining popularity. Here, too, I will claim that views of this kind fall prey to an objection from which self-deception as omission escapes, but in this case, the objection has not, to my knowledge, been made by others.<sup>15</sup> The strategy that I have in mind is to treat self-deception as a kind of pretense. This is a view of type (1) since it involves claims that the self-deceptive state is not a belief state. I will focus on Stephen Darwall's (1988) version of the view, but Tamar Gendler (2007) has also more recently proposed a similar view, and Jason D'Cruz (InPrep) revises Gendler's view. According to the self-deception as pretense view, when one is self-deceived, one is engaged in an elaborate pretense according to which *p* is the case. One acts as if *p* were true, but unlike in ordinary pretense where one also believes that one is engaged in pretense, when one is self-deceived, one is also engaged in a second-order pretense about one's first-order pretense: one behaves as if one is not merely behaving as if *p*. As Darwall (1988) puts it, this is "not simply the first-order pretense involved in fantasy, but also the second-order pretense that ... pretensions are real. When the self-deceiver plays the role of fool to himself, he must also pretend that he is not playing that role" (pp. 414–415).

This kind of account displays a sensitivity to the potentially distorting effects of motivation and affect, and makes an interesting appeal to the idea of fantasy which resonates with a lot of our ordinary experiences of self-deception. However, there is a problem. I take it that the reason that the second-order deception is necessary is to conceal from the self-deceiver the fact that she is engaged in pretense. Now, we have to face squarely the question of how someone could ever manage to get herself into that state in the first place. Plausibly, the purpose of engaging in the pretense is to preserve one's self-image or to avoid facing up to some painful facts. But of course this can't be achieved by pretense if one knows that one is pretending. The purpose of engaging in the pretense will be something which cannot be achieved unless it is hidden from the agent. That is the purpose of the second-order pretense.

However, if the reason for engaging in the second-order pretense is to make the first-order pretense more credible (i.e., to conceal it as pretense), how are we to make sense of the act of engaging in that second-order pretense without attributing to the agent the very knowledge that would undermine its aim? It seems that the agent must intend to get herself into the state of engaging in both pretenses for the sake of achieving a psychological end, but she must somehow manage to do this without revealing to herself that this is what she is doing.

One way of bringing out this difficulty is to ask why Darwall thinks second-order deception will be enough to do the trick. If the reason for engaging in the second-order pretense is to make the first-order pretense more credible, then wouldn't we also require a third-order pretense to conceal the nature of the second-order pretense as pretense, and so on, seemingly indefinitely? The self-deception-as-pretense strategy seems just to

harbor the very same dynamic problem as the naïve view in a somewhat concealed form – the dynamic problem becomes this: how can I act in a motivated and purposive way to get myself into a state of pretense which I do not know is a pretense?<sup>16</sup> The heart of the problem here is that self-deception requires a degree of doxastic engagement that mere pretense, if it is to remain distinct from belief, cannot sustain. However, if we allow pretense to have the required degree of doxastic engagement, it will be just as puzzling how one could intentionally achieve such pretense as how one could intentionally believe. Self-deception as omission escapes this objection by separating the process of coming into the nascently self-deceptive state from the episode of mental agency that makes that state into one of genuine self-deception. Indeed, since there is nothing mysterious about how one gets into the nascently self-deceptive belief state according to my view, the basic motivation for type (1) views – to give a non-doxastic account of the self-deceptive state in order to avoid the dynamic problem – is also undercut.

#### 4. Conclusion

In this paper, I introduce my view of self-deception and argued that it both avoids the dynamic problem of self-deception and captures the responsibility of self-deceivers. I also respond to Levy's argument that self-deceivers are not generally responsible, and argue that the responsibility of self-deceivers should be understood as that attaching to a voluntary failure to bring one's beliefs into conformity with epistemic requirements. I distinguish my view from other views on offer in the literature and argue that it has significant advantages over them by being immune to objections – old and new – which other views fall prey. Self-deception as omission is sensitive to how many of our mental processes can set us up for rational failings down the road. The fact that our minds are fragmentary and have inbuilt tendencies to bias is a source of moral and epistemic vulnerability against which the norms of epistemic conduct are meant to serve as a bulwark. Self-deceivers fail to guard as they ought to against these vulnerabilities.

#### Notes

1. The most puzzling feature of the naïve account is not that it requires self-deceivers to have contradictory beliefs. There might be something strange about believing the conjunction of  $p$  and not- $p$ , but nothing at all prevents an agent from believing  $p$  and believing not- $p$  at the same time. Indeed, it is almost certain that the body of my beliefs contains a pair such as this.
2. This is, of course, only a problem if self-deception is indeed possible. I assume throughout that it is possible because it is actual, and that skepticism about self-deception, thus, ought to be avoided.

3. My view is thus stated so as to give a merely sufficient condition for self-deception.
4. To add a word about the sense in which the evidence against the belief must be “available,” it will almost certainly be false that the agent “has” the evidence, in the sense of already having apprehended and appreciated it, however, in cases of genuine self-deception, it will be true that the evidence is there in the world, ripe for discovery. Self-deception is, among other things, to believe, for motivationally biased reasons, *in the face of evidence* to the contrary. It is an interesting consequence of thinking of self-deception in this way that the difference between self-deception and wishful thinking becomes a matter of degree determined by the strength of the evidence which is available. Intuitively, where self-deception is believing in the face of evidence to the contrary, wishful thinking requires only belief in spite of a lack of evidence in support. If this account is correct, however, the quality and abundance of the evidence available to the agent can vary independently of whether the belief-formation and the corresponding voluntary omission that are distinctive of self-deception have been deployed or perpetrated. Two agents who have the same beliefs can be such that one is a self-deceiver and the other is merely a wishful thinker if the former is in a situation where the available evidence speaks strongly against the belief she has, and the latter is in a situation where the evidence is weak or equivocal.
5. This is similar to how I state the view elsewhere in a different connection. See Gibson (2017). For elaboration on what it means for the agent’s reasons for omitting to seek, recognize, or appreciate evidence to derive from her desire that *p* be true see: Gibson (2017, p.4).
6. Sometimes, the term ‘dual-process theory’ is used to distinguish any model of a cognitive process which posits at least two independent streams of processing, leading to a single outcome. However, the term is also often used synonymously with ‘dual-systems theory,’ and that is the use that I adopt here.
7. I do not need the claim that all S1 processes involve propositional attitudes, just the significantly weaker claim that some S1 processes produce beliefs.
8. A methodological remark is in order here. I should thank an anonymous reviewer for reminding me that intuitions will diverge on cases such as these. When this happens, there may be no way, even in principle, to decide whose intuitions ought to serve as a guide. My intuitions, though they have certainly been tutored by the exercise of writing this paper, say that the gambler and the detective are self-deceived when they have the appropriate motivation. However, I do not think they should count as self-deceived only because of my intuitions. I have tried to begin with what I take to be the fairly firm judgment that self-deception is a species of a kind; the kind is, roughly, blameworthy epistemic failure. I have proposed features that distinguish self-deception from other species of this kind. Neither my judgment about species, nor my judgment about kind are beyond reproach, but it will take more than conflicting intuitions to dislodge them. For example, if one has the intuition that the motivated detective is not self-deceived, we either need to be told what other kind of blameworthy epistemic failure she is engaged in, or why it is not an instance of that kind, and why the line should be drawn at one of those two places instead of where I have proposed it be drawn.
9. My view, thus, has close affinities with Nelkin’s (2012) view of responsibility for self-deception.
10. I am of the view that epistemic vices are important objects of study for epistemologists. For a recent call to study epistemic vices more closely, see Cassam (2016).
11. This is, in fact, far from clearly the case. See further.

12. At this point, if Levy would be inclined to respond by claiming that nothing other than doubt could serve as the basis for the activation of System 2, my response would be simply to point out that System 2 can be activated for all sorts of reasons. Even if the effect of System 2 activation is “checking one’s work,” that doesn’t entail that it was for the sake of checking one’s work that it was activated.
13. Nelkin (2012, p. 135) also claims (and I am inclined to agree) that belief is compatible with doubt in a stronger sense, being not-completely-certain.
14. Perhaps most famously, this kind of view was defended by Davidson (2004a, 2004b). A version has also been defended by David Pears (1984).
15. It is made, however, much more briefly in Gibson (2017), and is inspired by Lockie (2003).
16. A version of this objection seems to work as well on D’Cruz’s account, according to which the pretense involved in self-deception is somehow unwitting. At first, it might seem that understanding self-deception as unwitting pretense might seem to help with this difficulty a little bit. After all, since the pretense is unwitting, perhaps it comes, as it were, with its aims already concealed, obviating the need to posit an ever-higher-order state of pretense to do the concealing for us. However, I worry that understanding unwitting pretense as the kind of metacognitive failure D’Cruz appeals to (i.e., failure to keep track of which of one’s pretenses are pretenses) might just reinstate the dynamic problem in a different form. It may be true that once one has lost track of whether one is pretending, the pretense will no longer be transparent in the way that it is if it is witting. However, again, how does one intentionally get into that state? Can one, in a motivated and purposive way, make it the case that a particular meta-representational content is only intermittently available? Why is this any less puzzling than trying to act in full knowledge to conceal as pretense a pretense whose aim can only succeed if one is unaware that it is a pretense?

## Disclosure statement

No potential conflict of interest was reported by the author.

## ORCID

Quinn Hiroshi Gibson  <http://orcid.org/0000-0002-9584-7049>

## References

- Cassam, Q. (2016). Vice epistemology. *The Monist*, 99(2), 159–180. <https://doi.org/10.1093/monist/onv034>
- Clifford, W. (1999). The ethics of belief. In T. J. Madigan (Ed.), *The ethics of belief and other essays* (pp. 70–96). Prometheus Books.
- D’Cruz, J. (InPrep). ‘Unwitting pretense and the self-deceptive mind’.
- Darwall, S. (1988). Self-deception, autonomy, and moral constitution. In B. Mclaughlin & A. O. Rorty (Eds.), *Perspectives on self-deception* (pp. 407–430). University of California Press.
- Davidson, D. (2004a). Deception and division. In *Problems of rationality* (pp. 200–212). Oxford University Press.

- Davidson, D. (2004b). Paradoxes of irrationality. In *Problems of rationality* (pp. 169–188). Oxford University Press.
- Ditto, P., & Lopez, D. (1992). Motivated skepticism: Use of differential decision criteria for preferred and nonpreferred conclusions. *Journal of Personality and Social Psychology*, 63(4), 568–584. <https://doi.org/10.1037/0022-3514.63.4.568>
- Epley, N., & Gilovich, T. (2016). The mechanics of motivated reasoning. *The Journal of Economic Perspectives*, 30(3), 133–140. <https://doi.org/10.1257/jep.30.3.133>
- Gendler, T. (2007). Self-deception as pretense. *Philosophical Perspectives*, 21(1), 231–258. <https://doi.org/10.1111/j.1520-8583.2007.00127.x>
- Gibson, Q. H. (2017). Self-deception in and out of illness: are some subjects responsible for their delusions?. *Palgrave Communications*, 3(15), 1–12. doi: doi.10.1057/s41599-017-0017-0
- Ginossar, Z., & Trope, Y. (1987). Problem solving in judgment under uncertainty. *Journal of Personality and Social Psychology*, 52(3), 464–474. <https://doi.org/10.1037/0022-3514.52.3.464>
- Johnston, M. (1988). Self-deception and the nature of mind. In B. McLaughlin & A. O. Rorty (Eds.), *Perspectives on self-deception* (pp. 63–91). University of California Press.
- Kahneman, D. (2011). *Thinking, fast and slow*. Farrar, Straus, and Giroux.
- Kahneman, D., & Tversky, A. (1972). On prediction and judgment. *Oregon Research Institute Bulletin*, 12(4).
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480–498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Levy, N. (2004). Self-deception and moral responsibility. *Ratio*, 17(3), 294–311. <https://doi.org/10.1111/j.0034-0006.2004.00255.x>
- Lockie, R. (2003). Depth psychology and self-deception. *Philosophical Psychology*, 16(1), 127–148. <https://doi.org/10.1080/0951508032000067707>
- Mele, A. (1997). Real self-deception. *Behavioral and Brain Sciences*, 20(1), 91–136. <https://doi.org/10.1017/S0140525X97000034>
- Mele, A. (2001). *Self-deception unmasked*. Princeton University Press.
- Nelkin, D. (2012). Responsibility and self-deception: A framework. In *Humana.Mente*, 4(20), 117–139.
- Pears, D. (1984). *Motivated irrationality*. Oxford University Press.
- Santioso, R., Kunda, Z., & Fong, G. T. (1990). Motivated recruitment of autobiographical memory. *Journal of Personality and Social Psychology*, 59(2), 229–241. <https://doi.org/10.1037/0022-3514.59.2.229>
- Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, 5(1), 207–233. [https://doi.org/10.1016/0010-0285\(73\)90033-9](https://doi.org/10.1016/0010-0285(73)90033-9)