

The Relationship between Self-Deception and Other-Deception

Anna Wehofsits

Preprint, the final article is available open access at:

The Southern Journal of Philosophy (2023)

<https://onlinelibrary.wiley.com/doi/full/10.1111/sjp.12540>

Abstract: Unlike the question of whether self-deception can be understood on the model of other-deception, the relationship between the two phenomena at the level of practice is hardly ever explored. Other-deception can support self-deception and vice versa. Self-deception often affects not only the beliefs and behavior of the self-deceiving person but also the beliefs and behavior of others who may become accomplices of self-deception. As I will show, however, it is difficult to describe this supportive relationship between self-deception and the deception of others without conceptual contradiction. While “deflationary” approaches offer a convincing way to avoid the so-called paradoxes of self-deception, they do not resolve the conceptual tensions that arise here. I conclude by outlining a solution.

1. Introduction

Reports on so-called fake news and alternative facts are pervasive. Opposing camps loudly accuse each other of deception and denial, be it with regard to climate change, police violence or Covid-19. Often the opponents defend even extremely far-fetched positions with such vehemence that it seems reasonable to assume that they believe them themselves. Complicated interactions of self-deception, other-deception and other cognitive and communicative failures seem to distort both individual and collective processes of belief formation.

In recent philosophical debates, the psychological and social interactions of self-deception and other-deception play hardly any role. They explore the relationship between the two phenomena almost exclusively with regard to the question of whether self-deception can be thought without contradiction when it is understood in analogy with other-deception. According to this analogy, deceiving oneself essentially has the same structure as deceiving others: one intentionally communicates something that one believes to be false (or at least not exactly true). The only significant difference is that in self-deception one is not deceiving others but oneself. This difference may sound harmless, but it is not. It takes complicated and controversial assumptions about the nature of the human mind to avoid that deceiving

oneself appears paradoxical. For this reason, most authors today defend a “deflationary” account and argue that we should conceptually separate self-deception from other-deception. I generally share this view. What is usually overlooked, however, is that the practice of self-deception maintains a relationship with other-deception that in turn also raises interesting conceptual and normative questions. Self-deception typically is a process with social, interactive components. Very often it affects not only the beliefs and behavior of the self-deceiving person but also the beliefs and behavior of others who may, voluntarily or involuntarily, become accomplices in self-deception. Drawing on everyday experience and psychological studies, it is plausible to assume that the deception of others can have a supporting function for self-deception and vice versa. Both in terms of how deception emerges and of how it is maintained, the two seem to interact dynamically. It is, however, very difficult to describe the relationship between self-deception and other-deception without conceptual contradiction. In what follows, I will first explain in general terms how I understand both phenomena and then discuss the problems associated with different views of their mutual support. I will show that deflationary proposals that help to resolve the tensions *within* the notion of self-deception do not also resolve the tensions that arise *between* self-deception and other-deception when we try to explain how they support each other. Finally, I will briefly outline how deflationary approaches can address this challenge and how they can analyze the relationship between self-deception and other-deception.

2. Conditions of other-deception and self-deception

To work out the tensions I am concerned with, it is, fortunately, not necessary to give a comprehensive definition of the heterogeneous phenomena of the deception of others and self-deception (if that is even possible). Rather, it is sufficient to address a few general conditions that apply to a wide range of manifestations of both. In accordance with the majority of historical and current positions, we can characterize the deception of others by at least two necessary conditions.¹ The first condition is that a communicative act² can only be a deception of others if the deceiving person convinces another person of something that she at the time of the communicative act believes to be false or to be something of whose

¹ For an overview of the different positions and the debate on possible further necessary conditions, see Mahon (2016).

² An act of deceiving others can of course involve several sub-acts and extend over a longer period.

correctness she is not convinced. More precisely, a communicative act can only be a deception of others if the deceiving person causes another person to adopt (or maintain) a belief that the deceiving person at the time of the communicative act believes to be false or to be something of whose correctness she is not convinced by providing what she believes to be fake or insufficient evidence on the basis of which the deceived person adopts (or maintains) the belief in question. I call this condition the *condition of conscious synchronic discrepancy*: by means of supposedly fake or insufficient evidence, the deceiving person causes another person to adopt or maintain a belief that deviates from her own beliefs at the time of the deception, and she is aware of this deviation.

The second condition for other-deception is that the deceiving person must have the intention to deceive.³ Most authors understand the intention to deceive as a conscious intention. Following Anscombe, they assume that intentional action involves some sort of awareness of what one is doing.⁴ Intentions thus form a subset of the broader category of motive, which includes conscious and unconscious motives.

It has traditionally⁵ been assumed that these two conditions also apply to self-deception.⁶ However, since in the case of self-deception, the deceiving and the deceived party coincide in one person, applying the two conditions there leads to considerable conceptual tensions.

³ In everyday language, we sometimes use the term "deception" in a broader sense to cover cases of inadvertently or mistakenly causing false beliefs, as may be the case, for example, with ironic or fictitious speech when the recipients do not recognize it as such and therefore misunderstand it. In the context of a philosophical inquiry, however, it is useful to make finer conceptual distinctions. For example, we may refer to cases of unintentionally causing false beliefs as unintentional "misleading" (Carson 2010, p. 47, see also Mahon 2016). As there are important differences between intentional deception and unintentional misleading - particularly with respect to active participation in and responsibility for causing false beliefs - I reserve the term "other-deception" (or "deception of others") in the following for intentional deceptions and exclude unintentional misleading. Since the attempt to deceive others proceeds indirectly, through evidence, it can be distinguished from other forms of intentionally manipulating other people's beliefs that fully bypass their agency, such as neural manipulation (see Mahon 2007, p. 185). Deceiving others can also be successful if the deceived person only implicitly adopts or maintains the belief in question, which may show itself in her behavior.

⁴ Anscombe (1957).

⁵ For a good summary of the traditional view, see Deweese-Boyd (2017).

⁶ I restrict myself here to these two necessary conditions because their transfer from other-deception to self-deception leads to conceptual tensions that deflationary approaches try to resolve by rejecting the two conditions for self-deception (see below). In contrast, deflationary approaches generally agree with traditional approaches that a third condition is necessary for both deception and self-deception: (self-)deception causes the deceived person to have a false belief or at least to be epistemically worse off. Since this third condition does not affect the conceptual tensions I am concerned with, I will not discuss it in what follows.

The attempt to transfer the condition of conscious synchronic discrepancy leads to a tension because it then implies that the self-deceiving person adopts or maintains a belief of which she is at the same time aware that it contradicts other of her beliefs. Thus, she should be able to judge: "I believe here and now that both p and not- p are true." But such a judgment is incomprehensible and makes self-deception appear to be a "static paradox," to use Alfred Mele's famous expression.⁷ According to my understanding of self-deception, the paradox does not necessarily concern the first-order content of the beliefs in question (such as "I am open-minded" [p] and "I am not open-minded" [not- p]). Rather, it may also arise from conflicting degrees of conviction regarding these contents, that is, from conflicting second-order beliefs (for example, "I firmly believe in being open-minded" [p] and "I do not firmly believe in being open-minded" [not- p]).

The attempt to transfer the condition of a conscious intention to self-deception, too, leads to tensions. How can a person consciously deceive herself into adopting or maintaining a belief that she believes to be false? Conscious intentions to deceive seem to undermine themselves as soon as they aim at the deceiving person herself: when we, as addressees of the deception, know about the intention to deceive, it will hardly be successful. In the current debate, this difficulty is called the "dynamic paradox." In my view, however, the "dynamic paradox" is not a real, logical paradox but only a psychological improbability. I believe a conscious intention to deceive can be successful and lead to self-deception, even if that is not common and does not represent a paradigmatic case of deceiving oneself.⁸

Because of these tensions, most authors today think that the two conditions apply to other-deception but not to self-deception. I share this view and believe that, on a general and abstract level, a (moderately) deflationary, inclusive approach best accommodates the different facets of self-deception.⁹ Depending on the context, the different facets can then

⁷ See Mele (1987; 2001). Mele has played a major role in shaping the current discussion about different variants of deflationary approaches. He also introduced the term "dynamic paradox" (see in the next paragraph) into the debate.

⁸ I agree with Mark Johnston (1988) that by using "autonomous means" we may succeed in intentionally adopting a belief that we do not currently hold. Pascal's famous wager can serve as an example: since one cannot simply believe in God at will, he recommends that to overcome their unbelief unbelievers act as if they believed in God, see Pascal (1910).

⁹ For an overview and discussion of different variants of deflationism, see Funkhouser (2019, pp. 58–134). However, some approaches that Funkhouser describes as revisionist (pp. 171–202) can also be classified as deflationary if, like Kevin Lynch (2017), one distinguishes between agential and non-agential forms of deflationism. Rejecting the condition of a conscious intention does not imply that

be further specified, and literary examples can help us understand why some of them count as deceptive in a stronger sense than others (Tolstoy's *Anna Karenina*, for example, can be read as an anthology of different manifestations of self-deception). Yet, as I explain in what follows, even with a general definition that does without such specifications, it is very difficult to conceptualize the relationship between self-deception and other-deception in a non-contradictory way.

Generally speaking, a person is deceiving herself about *p* when, due to her biased, distorting treatment of evidence that, given her other beliefs, better supports not-*p*, she believes that *p* (or has a somewhat weaker doxastic attitude towards *p*). This biased treatment is motivated by a directional attitude (possibly containing several components) she is (usually) not clearly aware of, such as a specific desire, a more general need, or a feeling of anxiety.¹⁰ To put it in less technical terms: in situations in which, without a distorting motive, the person would interpret the available evidence in a certain way (for instance, as clear evidence of cheating in a relationship), she blocks that interpretation self-deceptively by diverting her attention or interpreting the evidence in some other, pseudo-rational way. The distortion thus is not caused by arbitrary influencing factors such as fatigue or misleading evidence. Rather, it is motivated by some directional attitude, such as fear of the consequences the cheating might have (for example, fear of being alone or losing material security), shame at

the self-deceiver cannot be actively involved in her self-deception. Deflationists deny that self-deceptive beliefs are (typically) caused by an intention *to deceive*, but they may well believe that intentional actions are involved in the formation of self-deceptive beliefs. Deflationism does not imply a rejection of agency and responsibility. Some deflationists, like Neil Levy (2004), believe that we are not responsible for paradigmatic cases of self-deception. Other, more moderate deflationists, including me, believe that typical processes of self-deception contain not only uncontrolled but also controlled or at least (indirectly) controllable components on whose basis a person can be responsible for her self-deception. Psychological studies indicate that under certain conditions we are able to control even automatic or unconscious psychological processes. For instance, people who are committed to responding without bias manifested significantly less negative racial bias in a series of tests for implicit bias, see Devine and Plant (2002). And even if people cannot avoid responding with certain biases in certain situations, the decision not to be guided by these biases gives them considerable control over how these biases influence their further thinking and acting, see Bargh (1994). Thus, lack of control at the beginning of a process does not imply that its further course, too, is uncontrollable.

¹⁰ With the term "directional attitude" I take up Ziva Kunda's famous distinction between accuracy goals and directional goals: "accuracy goals lead to the use of those beliefs and strategies that are considered most appropriate, whereas directional goals lead to the use of those that are considered most likely to yield the desired conclusion." Kunda (1990, p. 481). Complex combinations of both types of goals can occur.

having been cheated on, the desire to keep the family intact, or a tendency to avoid conflict. As these examples illustrate, self-deception is often a reactive phenomenon, a defensive mechanism by which we respond to certain threats we perceive more or less vaguely,¹¹ but it can also serve more generally to prevent possible threats.

This general definition is deflationary in the sense that it avoids the paradoxes laid out above by setting weaker conditions for self-deception than the traditional understanding: it drops the condition of a conscious conflict of synchronic beliefs and merely presupposes a biased, distorting treatment of the available evidence.¹² Moreover, it replaces the condition of a conscious intention to deceive with the broader and weaker condition of a motive that the person deceiving herself does not need to be aware of. It only requires that she has some directional attitude that motivates her distorting treatment of the evidence. Furthermore, the definition is inclusive because it allows for various directional attitudes as motives and beliefs or belief-like attitudes as results of self-deception. It is therefore compatible with a number of deflationary approaches, such that the following considerations are relevant to them all.

As much as I think it is right to disconnect the concept of self-deception from the model of other-deception, it seems important to me not to overlook that in practical terms there is an important supportive connection between the two, which is not often addressed and even less often analyzed in detail. Recent philosophical debate has focused primarily on *intrapsychic* aspects of self-deception. Few authors address the social conditions that accompany the *process* of deceiving oneself, and hardly anyone investigates how social, *interpsychic* aspects contribute to the emergence and maintenance of self-deception.¹³ As

¹¹ See Holton (2022).

¹² This leaves open the possibility that self-deceivers have *diachronically* contradictory beliefs and are aware of them. It also leaves open the possibility that they *synchronously* have contradictory beliefs of which they are *not* aware. Presumably, we all have *synchronously* contradictory beliefs whose contradictions we do not notice because they are not obvious (as can happen when the contradiction results from complicated or indirect inferential relationships between the beliefs in question).

¹³ Exceptions include Beier (2010), Dietz (2017), Dings (2017), Galeotti (2018), Martin (2009), Oksenberg Rorty (1994), and Ruddick (1988). Amélie Oksenberg Rorty explicitly points out that people who are deceived by others participate in their deception and that people who deceive themselves depend on the complicity of others. But I think she goes too far when she claims that therefore the distinction between deceiving others and deceiving oneself is arbitrary, see Oksenberg Rorty (1994, p. 226). I agree that in concrete cases a clear distinction can be difficult or even impossible. Nevertheless, a conceptual distinction is possible and useful in understanding how the two phenomena interact.

William Ruddick notes as early as 1988, the philosophical discussion of self-deception has a “serious individualistic bias” that continues to obstruct our view of the social dimensions of self-deception. By way of contrast, he quotes Tolstoy, who uses the example of impatient relatives to show in just one laconic sentence how complicated the social dependencies of self-deception can be: “They all had only one wish that he [Nikolai] would die quickly, and they all did their best to conceal it and went on giving him medicines out of bottles, tried to discover new remedies and doctors, and deceived him and themselves and one another.”¹⁴ Similarly, Kant, too, already sees a connection between self-deception and the deception of others. On his view, longer processes of self-deception especially lead to the deception of others: “[the depravity of human nature] is characterized by a certain perfidy on the part of the human heart (*dolus malus*) in deceiving itself as regards its own good or evil disposition [. . .]. This dishonesty, by which we throw dust in our own eyes and which hinders the establishment in us of a genuine moral disposition, then *extends itself also externally to falsity [and] deception of others.*”¹⁵

Kant is concerned here with a specific form of self-deception, namely self-deception regarding one’s own moral integrity. Psychological studies as well as everyday experience suggest that the desire to perceive one’s behavior as morally justified is a very common motive for self-deception. Generally, self-deception is often about self-perception: it is often a more or less conscious, defensive strategy, a protective response to threats to how we want to perceive ourselves. Of course, it can also refer to other people and to events, but even then it is often about people or events that affect our self-perception. A psychological study by Wenger and Fowers with the telling title “Positive Illusions in Parenting: Every Child is Above Average,” for instance, shows that parents often deceive themselves about their children, which arguably reveals less about the children than about the parents themselves.¹⁶ Especially in the short term, self-deception can increase our psychological stability. In the longer term, however, it is often precarious because it can lead to bad practical decisions, painful disillusionment, and social isolation.

Kant’s moral-psychological thesis that a process that begins as self-deception eventually “extends” externally to deceiving others allows for different interpretations. My aim in the

¹⁴ Tolstoy (1961, p. 287), see Ruddick (1988, pp. 380, 383).

¹⁵ Kant (2018, AA 6:38), my emphasis.

¹⁶ Wenger and Fowers (2008).

following sections is not to reconstruct Kant's position. In fact, Kant does not elaborate on how he conceives of the relationship between self-deception and other-deception - perhaps because he did not notice any conceptual tensions here (whereas he explicitly mentions the tensions within the concept of self-deception that are at the center of the debate about intentionalist and deflationary conceptions of self-deception today¹⁷). Instead, I focus on three exemplary and at first sight plausible hypotheses about how self-deception can be "extended" to the deception of others and show that, on closer examination, they all prove to be problematic.

3. First reading of the extension thesis: other-deception as additional deception in consequence of long-term, stable self-deception

A first possible reading of the extension thesis is this: especially longer processes of self-deception, if they are successful and result in stable self-deception, lead (perhaps not always, but often) to a deception of others in the sense of an additional deception. The idea here is that the self-deceiving person is not only deceiving herself but eventually others as well. Successful, stable self-deception means that the person no longer has distressing doubts about her distorted beliefs. By avoiding, suppressing, or rationalizing counterevidence and emerging doubts, she has largely succeeded in immunizing her beliefs against contestation.¹⁸ At first glance, such an extension of stable, entrenched self-deception to other-deception may seem unproblematic. To show that this is not the case, I would like to return to the two conditions that, according to deflationary views, apply to other-deception but not to self-deception. If the process of self-deception is successful in the sense that the person no longer has distressing doubts about her distorted beliefs, this implies that she has no other beliefs of which she is aware that they contradict her distorted beliefs (on the level of content or in their degree of conviction). This means, however, that the first condition for the deception of others, the condition of conscious synchronic discrepancy, is not fulfilled. Since deceiving

¹⁷ Kant holds a traditional, intentionalist conception of self-deception, but refrains from resolving the conceptual tensions he identifies (Kant 1900-, AA 6:430). It is therefore unclear how he would spell out the extension thesis and whether he could do so without conceptual contradictions. I discuss Kant's view of self-deception in relation to passions in (Wehofsits 2020).

¹⁸ Some authors refer to what I call stable self-deception (and thus classify as a subset of self-deception) as self-delusion; for a discussion of these terminological differences, see Funkhouser (2019, pp. 124 – 128).

others requires a conscious, intentional deviation from one's own beliefs, communicating distorted beliefs to others cannot be considered as deceiving others if the communicating person herself believes in her distorted beliefs and has no other beliefs of which she is aware that they contradict the former beliefs. A person who successfully deceives herself simply passes on her distorted beliefs when she communicates them to others. Or, as George Louis Costanza in *Seinfeld* puts it: "Jerry, just remember. It's not a lie... if you believe it."

Some may still want to speak of other-deception here, perhaps to capture the intuition that a person communicating her self-deceptively distorted beliefs is at least partially responsible for passing on false information. If, by contrast, she were merely mistaken, she would not be responsible, or at least not to the same extent. I share this intuition, but I think it must be justified in another way. If the person is (partly) responsible, her responsibility is derivative. In analogy to responsibility in cases of drunk driving, where the person is responsible for drinking, it derives from the fact that the person is (partly) responsible for the preceding process of self-deception and therefore also (partly) responsible for its foreseeable consequences. Of course, it requires detailed description to determine whether in a given case, the person is responsible, and if so, to what extent.¹⁹ My only concern here, however, is to make it plausible that responsibility for communicating a distorted belief to others does not automatically mean that such communication is a case of other-deception.

It follows from these considerations that sufficiently successful or entrenched self-deception excludes other-deception for conceptual reasons. Sufficiently successful or entrenched self-deception rather results in the person passing on her *own* distorted beliefs to others. But of course self-deception is not always successful. On the contrary, it is often a precarious, unstable process accompanied by doubts and conflicts. This brings me to the second reading of the extension thesis.

4. Second reading of the extension thesis: other-deception as a substitute in consequence of admitted self-deception

According to the second interpretation of the extension thesis, deception of others can be the result of unsuccessful or admitted self-deception. If, for example, I have to admit to myself that I am not as skilled as I self-deceptively thought I was, it may be important to me that others, at least, think of me as particularly skilled. It can be important to me that others

¹⁹ See footnote 9 above.

maintain an image of me of which I can no longer convince myself (possibly because I consciously or unconsciously hope to regain the beliefs in question at a later time²⁰). Since in such cases self-deception ends before other-deception begins, there is no conceptual tension. Strictly speaking, however, there is no “extension” either. Self-deception is not extended by other-deception but replaced by it. This means that the person concerned does not simultaneously deceive herself and others about the same or closely related facts. The second reading of the extension thesis thus fails.

5. Third reading of the extension thesis: other-deception as a supportive, stabilizing component of processes of self-deception

A third reading retains the idea of extension and claims that there is an instrumental, supportive relationship between self-deception and other-deception. Briefly put, it does not say: first, I deceive myself, and if that works (first reading) or no longer works (second reading), then I deceive others. Rather it says: I deceive myself, or try to deceive myself, and in order to do so successfully, I deceive others. On this view, other-deception is not a consequence of successful or admitted self-deception; it is part of the process of self-deception and has an important, supportive function.

There is much to suggest that the deception of others can have a supporting function for unstable processes of self-deception. First, it is generally the case that many, if not most, of our beliefs require social support. How we see the world and ourselves depends largely on how others react to us, how they interact with us, whether they accept and affirm our descriptions of the world and ourselves, or question and reject them. In particular, self-descriptions that refer to us in relation to others—such as the degree of our attractiveness or kindness—must be validated by the affirmation of others. Second, when it comes to beliefs that are important to us, we need reassurance from others on a psychological level. Most urgently, we need such psychological reassurance when we have doubts about these important beliefs. The assumption therefore suggests itself that unstable processes of self-deception are often accompanied by other-deception. After all, self-deception is often only worth the effort when it comes to beliefs that are important to us, and doubts are typical side

²⁰ This points to the more complex relationship between self-deception and other-deception that I discuss in the following sections.

effects of processes of self-deception. In the course of such processes, the deception of others can fulfil a double function: it can help to overcome both inner and outer resistances. It can serve to reassure and trick oneself and also to reduce the risk of being disillusioned by others, for example by preventing them from asking unpleasant questions, expressing irritation, or even directly accusing us of self-deception.

To avoid misunderstandings, let me clarify. As an alternative, or an addition, to the scenarios discussed here, it is also possible that deception *by others* supports self-deception. This often happens, for example, in relationships of dependence and subordination. When people in positions of power self-deceptively attribute certain positive qualities to themselves and try to reassure themselves through the assurances of their subordinates, the latter have a strong incentive to give this reassurance, even if they do not believe that their superior has these qualities and realize that they are helping their superior's self-deception succeed. Think, for example, of a despotic boss who implicitly signals to his employees that he wants to hear from them how brilliant he is. Where employees depend on the goodwill of their boss, to avoid disadvantage for themselves, they may feign approval and give their boss the reassurance he is looking for (and presumably the boss will look for reassurance from employees who are unlikely to object). The case shows how complicated the relationship of self-deception and other-deception can be: the boss tries to make others accomplices in his self-deception by intentionally deceiving or unintentionally misleading them. But if they see through this and only feign approval, deception by others contributes to the boss's self-deception (provided he does not realize that he is being deceived).

Deception *by others* can thus be an important support for self-deception. This is especially true for collective processes of self-deception, where episodes of deception of and by others, self-deception, and more innocent misconceptions interact and stabilize each other. Such interactions can damage the corrective mechanisms that are available to a collective when it communicates openly, sincerely, and sufficiently free of distortions, and can thereby help ideologies to take root. In the context of this paper, however, I am not concerned with the relationship between self-deception and deception *by others* but with the relationship between self-deception and deception *of others* when both are exercised *by the same person*,

as is the case in the passage from Kant quoted above. I am thus concerned with cases in which a person deceives herself and others about the same or closely related facts.²¹

If we cannot implicitly trust that others will confirm our self-deceptive beliefs on their own—for reasons of dependence or politeness, for instance—then it is more likely that they will confirm them (and thus give us the reassurance we hope for) if we, as Kant calls it, “throw dust” not only in our eyes but also in theirs and “extend” our self-deception by deceiving them. At first glance, this sounds clear and simple. Extending one’s self-deception by the deception of others seems to be one of several strategies a person can use to stabilize her self-deception. If she tries to convince herself that *p*, against the available evidence and in violation of standards of rationality that she herself (at least implicitly) acknowledges, she can try to avoid the confrontation with unwanted evidence or reinterpret it. She can try to find supporting evidence. But she can also try to create alleged evidence herself—for example, by influencing the views and behaviors of others through deception in such a way that they support her own misjudgment.

But again, the devil is in the details: influencing the views and behaviors of others—for example, trying to convince others that one is not an envious person—can only be a deception of others (as opposed to unintentional misleading) if one is convinced or at least suspects that one is in fact an envious person. Furthermore, one must have a conscious intention to deceive. However, the more clearly one knows that one is an envious person, and the more clearly one recognizes one’s intention to deceive—together with the underlying motive of self-deception—the less likely it is that deceiving others will actually contribute to stabilizing one’s self-deception. The more clearly a case of influencing others is a case of deceiving them (because it fulfils both the condition of conscious synchronic discrepancy and the condition of a conscious intention to deceive), the less likely it is that it will contribute to stabilizing self-deception. Conversely, the less clear it is whether a case of influencing others fulfils both conditions, the less clear it is whether it is a case of deceiving others at all. In other words, the more clearly influencing others does not fulfil both conditions, the more clearly it tends merely to pass on self-deceptively distorted beliefs and thus merely to display one’s own self-deception.

²¹ Self-deception usually occurs as a more or less extensive cluster of related beliefs, emotions and behavioral dispositions that need not be expressible in a single proposition.

The reason why it is so difficult to conceptualize the deception of others as a social strategy to stabilize one's self-deception is that it seems that its success essentially depends on the fact that the supporting other-deception is precisely not a clear case of intentionally deceiving others. Of course, there are other social strategies besides other-deception for influencing other people's attitudes and behaviors in such a way that their feedback can be used to support one's self-deception. Roy Dings draws a helpful distinction between situating and persuasive social strategies of self-deception.²² When a person in support of her self-deceptive belief(s) surrounds herself with like-minded people or avoids people who think differently, she is using a situating strategy. In contrast, her strategy is persuasive when she actively tries to make other people like-minded by withholding information from them (negative persuasion), unintentionally²³ misleading them, or intentionally deceiving them (positive persuasion). Dings rightly emphasizes that it is important to distinguish between these different social strategies of self-deception in order to enable the empirical investigation of different social mechanisms of self-deception and effective preventive measures, and to make appropriate moral judgments about the blameworthiness of the self-deceiver. However, Dings does not address the question of how other-deception – in the sense of other-deception with the intention to deceive others – can be a means to self-deception without raising conceptual tensions. In fact, it is worth noting that he illustrates the two cases of positive persuasion with just one example – the example of Eleanor trying to impress others by namedropping well-known authors in order to get positive feedback that allows her to form and maintain the self-deceptive belief that she is highly intellectual.²⁴ Thereby, he leaves open the question of whether this is a case of unintentional misleading or other-deception as a means to self-deception. Just as Dings, I believe that both cases are possible. But only the latter raises the conceptual difficulties I am concerned with in this paper. In this sense, our respective approaches are complementary. We share the aim to contribute to a better understanding of the social strategies of self-deception. But while his focus is on identifying a number of different social strategies of self-deception, I focus on one of them, namely other-deception with the intention to deceive others, and present three

²² Dings (2017), p. 17-19.

²³ "Unintentional" in the context of this paper means that there is no intention to deceive; it does not imply that the action in question is not intentional in any way.

²⁴ Dings (2017), p. 19.

intuitively plausible ways to understand the relationship between a deflationary conception of self-deception and other-deception before discarding them on conceptual grounds.²⁵ In section 7, I will outline a proposal that enables me to spell out cases like Eleanor's in such a way that other-deception can successfully serve as a means to self-deception. First, however, I will discuss the converse case – self-deception as a means to other-deception –, for it also raises conceptual problems.

6. Self-deception as a supportive, stabilizing component of processes of other-deception

None of the three readings, then, can explain how self-deception is extended and supported by the deception of others. What makes matters even more complicated is that conversely, self-deception can also have a supporting function in deceiving others, as psychological studies confirm. According to social psychologist von Hippel and evolutionary biologist Trivers, self-deception evolved to facilitate other-deception.²⁶ It is easier to convince others

²⁵ Analyzing the relationship between intentionalist conceptions of self-deception and other-deception would be a separate project.

²⁶ Anna Elisabetta Galeotti distances herself from this evolutionary thesis (Galeotti 2018, p. 90), but agrees that self-deception can have a supporting function for other-deception. She discusses three possible relationships between self-deception and other-deception (Galeotti's use of "other-deception" includes both other-deception with and without the intention to deceive others). (1) Other-deception as a by-product of self-deception; (2) self-deception as an ancillary to other-deception; (3) self-deception providing the justification for *explicit* other-deception (Galeotti 2018, pp. 91-99). In case (1), other-deception as by-product corresponds to unintentional misleading. It is a case of what I characterized in section 3 as the passing on of one's own self-deceptive beliefs to others. The self-deception is motivated (see Galeotti 2018, pp. 56, 79) but the resulting misleading is not (neither by the intention to deceive others, nor by any other motive). According to Galeotti, case (2) is "specific to the political domain". Self-deception "concerns the cover-up of political fiascoes" (p.93), whereby "the wish which sets the self-deceptive process in motion is precisely convincing others - Parliament, the people, and international audiences" (Galeotti 2018, p. 96). When this wish leads to self-deception, and subsequently to the misleading of Parliament, the people, or international audiences, we are dealing with a complex case of motivated but unintentional misleading (after all, the desire to convince others is not an intention to deceive them). Yet, as Galeotti points out, the process is often "intertwined with [...] ideological convictions, unexamined assumptions, straightforward lies, and cold mistakes, so as to make the process very muddy." (Galeotti 2018, p.96) Intentions to deceive others may thus have a role to play. In case (3), "explicit" other-deception corresponds to intentional other-deception. However, since Galeotti focuses here on cases in which other-deception and self-deception refer to different facts, the tensions I have discussed do not arise. The other-deception is about one fact (as an example, Galeotti cites Kennedy's lie about the secret deal with the Soviet Premier Khrushchev (Galeotti 2018, p.99)) whereas the self-deception is about another, namely the moral status of the other-deception, which it represents as justified, for example as a "noble lie" (Galeotti 2018, 98-99). In this case, self-deception does not serve to conceal other-deception from oneself (by now believing oneself what one has pretended to others with the intention of deceiving

when we believe what we say ourselves. Self-deception allows “people to avoid the cues to conscious deception that might reveal deceptive intent”.²⁷ Building on this, Chance and Norton argue that self-deception can be a non-conscious, adaptive strategy for deceiving—or lying to²⁸—others without being exposed and possibly punished. Self-deception can prevent the deceiver from showing non-verbal signs of guilt. It can also reduce the cognitive effort involved in deceiving others.²⁹ After all, a person who deceives others must safely separate and manage two inferential systems of beliefs, her own and the one she tries to generate in another person. This inverse support, according to which self-deception can serve to stabilize the deception of others, also seems plausible. Especially in the case of long-term deception of others, it is plausible that self-deception can bring relief and protect from exposure and possible retribution. However, this reverse support, too, is difficult to conceptualize clearly. What starts as deceiving others with a double administrative burden does not simply become a less costly deception of others when the deceiving person finally deceives herself as well. Rather, in the course of this process, the deception of others increasingly loses what makes it a deception of others, it dissolves. In order to successfully support the deception of others, the self-deception must be as solid as possible. However, the more it gains in strength, the more the deception of others tends to become a mere (though possibly derivatively culpable) passing on of one’s own distorted beliefs.

It fits this diagnosis that von Hippel and Trivers use two different concepts of “interpersonal deception”. If their concept of interpersonal deception were always the same, this would lead to contradictions. However, they do not explain this dual use and often do not even make clear which of the two concepts they are referring to or that they are switching from one to the other. “[C]onsciously mediated deception that could reveal their deceptive intent”³⁰

them). Rather, other-deception is acknowledged and self-deception serves to justify it morally. Galeotti's focus is different from Dings's or mine, but we share the aim of better understanding the social relationships between self-deception and other-deception or unintentional misleading. We discuss conceptual alternatives that are appropriate (or, because of inherent tensions, inappropriate) for describing different relationships between self-deception and other-deception or unintentional misleading, and thus provide a conceptual toolkit that can help investigate, distinguish, and evaluate concrete cases.

²⁷ Von Hippel and Trivers (2011, p.1), cf. Trivers (2011).

²⁸ Lying is here understood as intended deception by untruthful assertion (further differentiations are possible but not important for my argument here).

²⁹ Von Hippel and Trivers (2011), Chance and Norton (2015).

³⁰ Von Hippel and Trivers (2011, p. 4).

corresponds to the notion of other-deception used here, which fulfills the condition of conscious synchronic discrepancy and the condition of a conscious intention. However, according to von Hippel and Trivers interpersonal deception can also be unconscious, so that the two conditions are not (any longer) fulfilled: "To the degree that people can convince themselves that a deception is true or that their motives are beyond reproach, they are no longer in a position in which they must knowingly deceive others. Thus, *the central proposal of our evolutionary approach to self-deception is that by deceiving themselves, people can better deceive others, because they no longer emit the cues of consciously mediated deception that could reveal their deceptive intent.*"³¹

The tensions I have identified can explain why von Hippel and Trivers must use two different concepts of interpersonal deception in formulating their evolutionary thesis. They understand "*conscious* interpersonal deception" as the evolutionary driver of self-deception.³² In addition, they use the term to refer to specific acts of interpersonal deception *before* the deceiving person deceives herself to eliminate the cues that reveal her deceptive intent.³³ "*Unconscious* interpersonal deception," in contrast, refers to acts of "deceiving" others that are performed *in* a state of self-deception and therefore no longer show signs of deceptive intent and are cognitively less demanding than conscious deceptions. I have distinguished between other-deception and unintentional misleading to make clear that different types of cognitive processes are involved in the two cases.³⁴ With their often ambiguous use of the term "interpersonal deception," von Hippel and Trivers obscure this distinction. It is not only conceptually but also morally important: unintentional misleading, all else being equal, is morally less problematic than other-deception (though the charge of inadvertence may also be appropriate in some cases of unintentional misleading). Von Hippel's and Trivers's "non-conscious interpersonal deception" seems to correspond to

³¹ Von Hippel and Trivers (2011, p. 4).

³² "[S]elf-deception evolved to facilitate interpersonal deception by allowing people to avoid the cues to conscious deception that might reveal deceptive intent." Von Hippel and Trivers (2011, p. 1).

³³ In some cases, Von Hippel and Trivers seem to assume, we can avoid these cues by self-deception in the first place (see their discussion of self-enhancing biases, von Hippel and Trivers (2011, p. 5)). In such cases, the act of self-deception (and any possibly resulting misleading of others - von Hippel and Trivers would again speak of "deception" here) is not preceded by an act of conscious interpersonal deception.

³⁴ See footnote 3.

unintentional misleading.³⁵ In his book, Trivers (2011) uses an even broader notion of deception, including crypsis (the ability of an animal or a plant to avoid detection by enemies) in addition to other-deception and unintentional misleading, and even the "deceptive molecular techniques" that "selfish genetic elements use to overreproduce at the expense of other genes".³⁶ The tensions I have identified and the solutions I outline in the concluding section may, I hope, help to clarify von Hippel's and Trivers's argument.

7. Conclusion

With regard to how they emerge and are maintained, there seems to be a mutual, supportive relationship between self-deception and other-deception: the deception of others can support self-deception and vice versa. However, a conceptual analysis shows how difficult it is to explain this relationship without contradiction. A person who deceives herself cannot at the same time deceive others about the same or closely related facts. Conversely, a person who deceives others cannot at the same time deceive herself about the same or closely related facts. Whether self-deception supports the deception of others or vice versa, they can never both be fully developed at the same time. Moreover, it is difficult to clearly distinguish between the individual episodes of self-deception and other-deception and to explain how they merge. The deception of others, it seems, is able to support self-deception precisely when it is not or no longer recognizable as such (for instance, when the despotic boss does not have to consciously lie to his employees in order to get confirmation from them that he is brilliant or when he successfully suppresses and forgets that he only receives their confirmation because he has previously deceived them). Conversely, self-deception only seems to be able to support the deception of others when the other-deception dissolves and becomes a passing on of one's own distorted beliefs (who knows, maybe Karl May, a best-selling German author of adventure novels³⁷, at some point believed himself that he had fought with grizzly bears and understood more than twelve hundred languages and dialects). In a sense, this repeats the problems of the static and the dynamic paradox that arise in transferring the condition of conscious synchronic discrepancy and the condition of a

³⁵ The conclusion that "unconscious" implies "unintentional" seems permissible if one agrees with Anscombe (1957) that intentional action presupposes some kind of awareness (see above).

³⁶ Trivers (2011, p. 7).

³⁷ See Wollschläger (2004).

conscious intention to deceive onto the concept of self-deception. The solution, however, cannot be the same: an understanding of self-deception that dispenses with both conditions can resolve the tensions *within* the notion of self-deception, but it does not eliminate the tensions that arise *between* self-deception and the deception of others in describing their mutually supportive relationship.

Nonetheless, it would be hasty to conclude that we must deny the existence of the relationship in question or else assume that it is conceptually impossible for deflationists to analyze the supportive relationship between self-deception and other-deception. Of course, there may be cases in which what at first glance appears to be a relationship between self-deception and other-deception is actually a relationship between self-deception and a possibly motivated but unintentional manipulation of other people's attitudes and behaviors, such as unintentional misleading or unintentional withholding of information.³⁸ It may sometimes happen, then, that we call other-deception what should be called unintentional misleading or unintentional withholding of information, and that in such cases a conceptual substitution can resolve the tensions I have pointed out. As a consequence of such a conceptual substitution, we may realize that we also have to revise our moral assessment of the case in question. However, in other cases it may be empirically and morally inadequate to make such conceptual substitutions; for explanatory and for moral reasons we should clearly distinguish between unintentional manipulation and intentional other-deception. In closing, I would like to outline briefly that it is possible to defend the plausible intuition, supported by psychological studies, that self-deception and other-deception can be mutually supportive if we look more closely at how they influence each other. In a diachronic perspective, self-deception and other-deception can at least indirectly support each other if episodes of self-deception and other-deception take place at different times or if there are periods between them in which other cognitive processes take place, such as forgetting, vivid imagining or rehearsal of misinformation³⁹, that make transitions between them intelligible.

³⁸ In contrast to the concept of other-deception, the broader concept of manipulation does not presuppose a conscious intention to deceive, and the condition of conscious synchronic discrepancy does not apply either. As Kate Manne illustrates with compelling examples, we can manipulate others without consciously intending to do so (see Manne (2014)). Successful manipulation can lead to the manipulated person adopting beliefs that contradict the manipulator's beliefs, but the concept does not imply this.

³⁹ Von Hippel and Trivers (2011, p. 6).

For example, it is possible that a person who deceives herself in period t_1 that p has to admit this deception to herself, against her will, in period t_2 . In period t_3 , she may succeed at least in successfully deceiving others that p . If she then, in period t_4 , forgets that she was deceiving others, relevant feedback from the deceived can help her in period t_5 to once more self-deceptively believe that p . In this context, it is likely that the motive of self-deception at t_5 —such as a desire to be highly intellectual⁴⁰—is the same as at t_1 . The same motive can also explain why the admission of self-deception at t_2 is against the person's will, and why at t_3 she deceives others. Again, episodes of deceiving oneself alternate with episodes of deceiving others, they occur at different times and support each other only indirectly. It does, nonetheless, become clear that there is a motivational connection between the episodes of self-deception at t_1 and t_5 , which is mediated by episodes of deceiving others and forgetting. Looking at this connection, we can see what structure unstable, changing processes of self-deception can have. Thus, the conceptual tensions I have discussed do not show that there can be no supportive relationship between self-deception and other-deception. Rather, they show that this relationship is even more dynamic and complex than it first appears to be.⁴¹

References

Anscombe, G. E. M. 1957. *Intention* (Oxford: Basil Blackwell).

Bargh, John 1994. "The Four horsemen of Automaticity: Awareness, Intention, Efficiency, and Control in Social Cognition," in *Handbook of Social Cognition: Basic Processes*, ed. Robert S. Wyer and Thomas K. Srull (Hillsdale, NJ: Erlbaum), pp. 1-40.

Beier, Kathi 2010. *Selbsttäuschung* (Berlin and New York: De Gruyter).

Carson, Thomas L. 2010. *Lying and Deception: Theory and Practice* (Oxford: Oxford University Press).

Chance, Zoë and Norton, Michael I. 2015. "The what and why of self-deception," *Current Opinion in Psychology*, vol. 6, pp. 104–107.

⁴⁰ See Dings's example of Eleanor above.

⁴¹ I would like to thank Monika Betzler, Katja Crone, Ophelia Deroy, Anna Goppel, Jan Niklas Howe, Gerhard Ernst, Felix Koch, Andreas Müller, Christof Rapp, Peter Schaber, Neil Van Leeuwen, two anonymous referees, and all who discussed earlier versions of this paper with me at the Center for Advanced Studies at LMU Munich and in colloquia in Bern, Bochum, Erlangen, Munich, and Zurich.

Devine, Patricia G., E. Ashby Plant, David M. Amodio, Eddie Harmon-Jones, and Stephanie L. Vance 2002. "The regulation of explicit and implicit race bias: The role of motivations to respond without prejudice," *Journal of Personality and Social Psychology*, vol. 82, no. 5, pp. 835-848.

Deweese-Boyd, Ian 2017. "Self-Deception," in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta. <https://plato.stanford.edu/entries/self-deception/>

Dietz, Simone 2017. „Selbsttäuschung als sozialer Prozess,“ in: *Selbsttäuschung: Eine Herausforderung für Philosophie und Psychologie*, ed. Emil Angehrn and Joachim Küchenhoff (Weilerswist: Velbrück), pp. 223-239.

Dings, Roy 2017. „Social Strategies in self-deception,“ *New Ideas in Psychology* 47, pp. 16-23.

Funkhouser, Eric 2019. *Self-Deception* (London and New York: Routledge).

Galeotti, Anna Elisabetta 2018. *Political Self-Deception* (Cambridge: Cambridge University Press)

von Hippel, William, and Robert Trivers 2011. "The evolution and psychology of self-deception," *Behavioral and Brain Sciences*, vol. 34, no. 1, pp. 1-16.

Holton, Richard 2022. "Self-Deception and the Moral Self," in *The Oxford Handbook of Moral Psychology*, ed. Manuel Vargas and John M. Doris (Oxford: Oxford University Press).

Johnston, Mark 1988. "Self-deception and the nature of mind," in *Perspectives on Self-Deception*, edited by Brian P. McLaughlin and Amélie Oksenberg Rorty, 63–91. (Berkeley: University of California Press).

Kant, Immanuel (1900-). *Gesammelte Schriften, Akademie Ausgabe*. (Berlin: Walter de Gruyter and predecessors).

Kant, Immanuel 2018. *Religion within the Boundaries of Mere Reason* (Cambridge: Cambridge University Press).

Kunda, Ziva 1990. "The case for motivated reasoning," *Psychological Bulletin*, vol. 108, no. 3, pp. 480-498.

Levy, Neil 2004. "Self-Deception and Moral Responsibility," *Ratio*, vol. 17, pp. 294-311.

Lynch, Kevin 2017. "An Agentive Non-Intentionalist Theory of Self-Deception," *Canadian Journal of Philosophy*, vol. 47, pp. 779-798.

Mahon, James 2007. "A Definition of Deceiving," *International Journal of Applied Philosophy*, vol. 21, no. 2, pp. 181-194.

----- 2016. "The Definition of Lying and Deception," in *Stanford Encyclopedia of Philosophy*, ed. Edward N. Zalta

Manne, Kate 2014. "Non-Machiavellian Manipulation and the Opacity of Motive," in *Manipulation: Theory and Practice*, ed. Michael Weber and Christian Coons (Oxford: Oxford University Press), pp. 221-245.

Martin, Clancy (ed.) 2009. *The Philosophy of Deception* (Oxford: Oxford University Press).

Mele, Alfred 1987. "Recent Work on Self-Deception," *American Philosophical Quarterly*, vol. 24, no. 1, pp. 1-17.

----- 2001. *Self-Deception Unmasked* (Princeton: Princeton University Press).

Oksenberg Rorty, Amélie 1994. "User-Friendly Self-Deception," *Philosophy*, vol. 69, pp. 211-228.

Pascal, Blaise 1910. *Pensées (Thoughts)*. Trans. W. F. Trotter, (London: Dent).

Ruddick, William 1988. "Social self-deceptions," in *Perspectives on Self-Deception*, ed. Brian P. McLaughlin and Amélie Oksenberg Rorty (Oakland: University of California Press), pp. 380-389.

Tolstoy, Lev 1961. *Anna Karenina*. Trans. D. Magarshack, (New York: Signet).

Trivers, Robert 2011. *The Folly of Fools. The Logic of Deceit and Self-Deception in Human Life*. (New York: Basic Books).

Wehofsits, Anna 2020 (online) / 2023 (in print). "Passions: Kant's psychology of self-deception", in: *Inquiry* 66(6): 1184–208, <https://doi.org/10.1080/0020174X.2020.1801498>

Wenger, Andrew and Blaine J. Fowers 2008. "Positive illusions in parenting: Every child is above average," *Journal of Applied Social Psychology*, vol. 38, no. 3, pp. 611-634.

Wollschläger, Hans 2004. *Karl May: Grundriß eines gebrochenen Lebens* (Göttingen: Wallstein).