Les ateliers de l'éthique
The Ethics Forum

COSTLY FALSE BELIEFS: WHAT SELF-DECEPTION AND PRAGMATIC ENCROACHMENT CAN TELL US ABOUT THE RATIONALITY OF BELIEFS

Melanie Sarzano

Volume 13, numéro 2, été 2018

URI : https://id.erudit.org/iderudit/1059501ar
DOI : https://doi.org/10.7202/1059501ar

Aller au sommaire du numéro

Éditeur(s)

Centre de recherche en éthique (CRÉ)

ISSN

1718-9977 (numérique)

Découvrir la revue

Citer cet article

Sarzano, M. (2018). COSTLY FALSE BELIEFS: WHAT SELF-DECEPTION AND PRAGMATIC ENCROACHMENT CAN TELL US ABOUT THE RATIONALITY OF BELIEFS. *Les ateliers de l'éthique / The Ethics Forum*, *13*(2), 95–118. https://doi.org/10.7202/1059501ar

Résumé de l'article

Dans cet article, je compare les cas classiques de duperie de soi aux cas que l'on trouve dans les débats sur la question de l'empiètement pragmatique et défends l'idée selon laquelle ces deux types de cas peuvent être compris comme étant produits par un même mécanisme visant à éviter la formation de croyances fausses coûteuses. Cette comparaison nous mène naturellement à former un dilemme à propos de la rationalité des croyances. Le dilemme repose sur l'idée que bien que ce mécanisme mène à la formation de croyances irrationnelles dans les cas de duperie de soi, il ne semble pas affecter la rationa-lité du sujet dans les cas d'empiètement pragmatique : alors que les sujets autodupés sont irrationnels, les sujets décrits dans les cas d'empiètement pragmatique ne le sont pas. Pour résoudre ce dilemme sans rejeter les présupposés selon lesquels les croyances issues de la duperie de soi sont irrationnelles et que les cas sur lesquels repose l'empiètement pragmatique sont rationnels, je propose plusieurs hypothèses visant à expliquer cette différence, prouvant ainsi que ce dilemme n'est qu'apparent et que l'irrationalité de la duperie de soi ne peut uniquement dépendre de ce mécanisme sous l'influence de considérations pratiques.

COSTLY FALSE BELIEFS: WHAT SELF-DECEPTION AND PRAGMATIC ENCROACHMENT CAN TELL US ABOUT THE RATIONALITY OF BELIEFS

MELANIE SARZANO
PHD STUDENT, UNIVERSITY OF ZURICH

ABSTRACT:

In this paper, I compare cases of self-deception and cases of pragmatic encroachment and argue that confronting these cases generates a dilemma about rationality. This dilemma turns on the idea that subjects are motivated to avoid costly false beliefs, and that both cases of self-deception and cases of pragmatic encroachment are caused by an interest to avoid forming costly false beliefs. Even though both types of cases can be explained by the same belief-formation mechanism, only self-deceptive beliefs are irrational: the subjects depicted in high-stakes cases typically used in debates on pragmatic encroachment are, on the contrary, rational. If we find ourselves drawn to this dilemma, we are forced either to accept—against most views presented in the literature—that self-deception is rational or to accept that pragmatic encroachment is irrational. Assuming that both conclusions are undesirable, I argue that this dilemma can be solved. In order to solve this dilemma, I suggest and review several hypotheses aimed at explaining the difference in rationality between the two types of cases, the result of which being that the irrationality of self-deceptive beliefs does not entirely depend on their being formed via a motivationally biased process.

RÉSUMÉ :

Dans cet article, je compare les cas classiques de duperie de soi aux cas que l'on trouve dans les débats sur la question de l'empiètement pragmatique et défends l'idée selon laquelle ces deux types de cas peuvent être compris comme étant produits par un même mécanisme visant à éviter la formation de croyances fausses coûteuses. Cette comparaison nous mène naturellement à former un dilemme à propos de la rationalité des croyances. Le dilemme repose sur l'idée que bien que ce mécanisme mène à la formation de croyances irrationnelles dans les cas de duperie de soi, il ne semble pas affecter la rationa-lité du sujet dans les cas d'empiètement pragmatique : alors que les sujets autodupés sont irrationnels, les sujets décrits dans les cas d'empiètement pragmatique ne le sont pas. Pour résoudre ce dilemme sans rejeter les présupposés selon lesquels les croyances issues de la duperie de soi sont irrationnelles et que les cas sur lesquels repose l'empiètement pragmatique sont rationnels, je propose plusieurs hypothèses visant à expliquer cette différence, prouvant ainsi que ce dilemme n'est qu'apparent et que l'irrationalité de la duperie de soi ne peut uniquement dépendre de ce mécanisme sous l'influence de considérations pratiques.

0. INTRODUCTION

Sometimes, subjects hold irrational, motivated beliefs in the face of evidence to the contrary, mistreat the evidence at hand, and seem impervious to any evidence contradicting their desire. Such subjects are likely to be self-deceived.

In reaction to more traditional views, *motivationists* about self-deception have recently argued that self-deceptive beliefs can be understood as *motivationally biased beliefs*: they are the result of the subjects' motivation, such as a desire or an emotion, that biases their treatment of the evidence, leading them to form this irrational belief (Johnston, 1988; Barnes, 1997; Funkhouser, 2005; Scott-Kakures, 2002; 2012; Lazar, 1999; Mele, 1997; 1999; 2001; Nelkin, 2002).

In this paper, I argue that, understood this way, self-deception shares interesting similarities with high-stakes cases borrowed from the literature on pragmatic encroachment. Despite these striking similarities, these two types of cases have rarely, if ever, been compared in either part of the literature.[1] Although these two types of cases differ in many ways, I argue that they can both be explained by reference to the same underlying belief-formation mechanism: a mechanism characterized by the influence of practical factors—in particular, the influence of costly errors. While Mele (1997; 1999; 2001) refers to this mechanism to explain the distorted ways in which subjects come to form irrational, self-deceptive beliefs, the same mechanism can be invoked to explain the subject's epistemic behaviour in rational high-stakes cases.

My argument rests upon the assumption that self-deceptive beliefs generated by this type of belief-formation mechanism are irrational, whereas high-stakes cases typically aren't. If this mechanism really does participate in making the former, but not the latter, irrational, we then face a dilemma about the rationality of beliefs. According to this dilemma—or, as we will see, what I take to be an *apparent* dilemma—we are forced either to accept that self-deception is rational or to accept that high-stakes cases are irrational.[2] In order to solve this dilemma, I suggest and review several hypotheses to explain this difference in rationality, and I argue that the best way of articulating the difference is by drawing a line between two types of motivation, or two types of costs influencing the belief-formation process. If this line of thought is correct and if there really is a way of dissolving the dilemma, then the irrationality of self-deceptive beliefs, and of irrational beliefs in general, cannot be explained merely by the fact that the beliefs in question result from a motivationally biased process.

In section I, I introduce self-deception as presented in the motivationist framework. In section II, I turn to classical high-stakes cases borrowed from the literature on pragmatic encroachment and put forward the idea that, contrary to cases of self-deception, these cases are cases in which the subject's epistemic state is rational. In section III, I begin by showing how both sets of cases can in fact be explained by referring to the same type of belief-formation mechanism: Friedrich-Trope-Liberman (FTL) model of lay hypothesis testing. On this basis,

I argue that, if we assume that this belief-formation process is what causes these beliefs to be irrational, then we are led to a dilemma about the rationality of beliefs. Finally, in section IV, I suggest several hypotheses for explaining this difference in rationality and evaluate them in turn, finally putting forward a solution to the dilemma.

I. SELF-DECEPTION

Some beliefs are undoubtedly irrational. Self-deceptive beliefs are of this type, they are beliefs held in the face of evidence to the contrary. Such cases typically involve subjects refusing to believe that their unfaithful partner is cheating on them, parents resisting the fact that their children aren't as perfect as they take them to be, and lovers mistakenly believing, despite continuous rejection, that their love is requited. In all of these cases, subjects aren't merely hoping or wishing reality were different: they seem to strongly believe that reality is such, that it matches their desires and respond to evidence in a very biased way, despite the facts being clear to everyone but themselves.

There are two main families of theories about self-deception. On the classical, *intentionalist* view (Davidson, 1985; 2004; Pears, 1984; Rorty, 1988; Talbot, 1995), self-deception is understood literally, as a case of deception in which the deceiver and the deceived are one and the same person. This model is closely built on our common understanding of interpersonal deception, of cases in which a first subject intentionally tricks another into believing something false. On the intentionalist view, then, subjects are self-deceived only if they intentionally deceive themselves and, at least on most intentionalist accounts, hold contradictory beliefs (an initial belief that not-*p* and a self-deceived belief that *p*). This approach has famously led to intricate puzzles, for pretty obvious reasons; what can easily be applied to two distinct minds in cases of interpersonal deception becomes overly complicated and tricky when applied to a single mind. How does a single mind intentionally deceive itself, since it is probably aware of its own intention to do so? And how is it possible for a single subject to come to believe *p* when that subject already believes, and knows, that this belief is false?

In response to these thorny issues, *motivationists* (sometimes also called *non-intentionalists*) have recently argued that self-deception need not involve an intention or any contradictory beliefs. On the motivationist view, all that is required for a self-deceptive belief to be formed is a motivational factor influencing the subject's belief-formation process (Johnston, 1988; Barnes, 1997; Funkhouser, 2005; Scott-Kakures, 2002; 2012; Lazar, 1999; Mele, 2001; Nelkin, 2002). Here, it is the subjects' desire or emotion that biases their belief formation and influences their treatment of the evidence, thereby resulting in them holding a self-deceptive, irrational, belief. Because the intentionalist view is problematic in many aspects, I will here assume that the motivationist framework is correct and work with the following—fairly uncontroversial—characterization of self-deception primarily inspired by Alfred Mele's deflationary

account (1997; 1999; 2001). On Mele's account, a subject *S* is self-deceived in believing a proposition *p* if the following conditions obtain:

(a) *unwarranted belief*
S's belief that *p* is false (or at least, unwarranted).

(b) doxastic alternative[3]
S possesses, or has been in contact with, evidence supporting the belief that not-*p*.

(c) motivated belief
S's belief that *p* is the result of a motivationally biased process.

These conditions capture what is essential to most motivationist accounts of self-deception: (a) that the self-deceptive belief is unwarranted (i.e., that the subject *should not* in fact believe *p* given the evidence at hand but believes *p* nevertheless);[4] (b) that the subject has been presented, or has been in contact with, evidence supporting the exact opposite of what he or she currently believes (i.e., that, given the evidence, he or she should in fact believe not-*p*);[5] and (c) that *S*'s belief that *p* is a direct consequence of the subject's motivation that *p*. To illustrate this view, consider the following example.

> *Fernando*
> Fernando is a soft-hearted postdoc who happens to be madly in love with Steve. His fondness for Steve has grown into an obsessive love over the years and it is clear that he is now craving to see his love returned. The two men accidentally run into each other every few months at conferences and workshops, make small talk, and act politely. Steve has, so far, never shown any sign of mutual attraction and his small talk and polite behaviour only ever warranted Fernando to believe that Steve simply appreciates him as a colleague. In fact, it should even be clear to Fernando, after several refusals to have coffee together, or spend time together, that Steve isn't romantically interested in Fernando. Nevertheless, Fernando fallaciously interprets Steve's every little gesture and word as conclusive proof of their impending love affair.

In this example, (a) Fernando unwarrantedly believes that his love is requited, (b) when he should in fact believe the exact opposite (i.e., that Steve isn't romantically interested), and, what is more, (c) Fernando's belief is formed and maintained through a motivationally biased process that influences his interpretation of the evidence at hand. In other words, Fernando is self-deceived, and sadly Steve isn't about to declare his love to him.

As I said, it is central to motivationist accounts that the subject mistreats the evidence at hand (cf. condition [c]) or is insensitive to evidence. According to Mele (1997; 1999; 2001), there are several ways in which a subject can mistreat evidence. In the aforementioned case, Fernando is mistreating the evidence,

mainly by ignoring the evidence against Steve being romantically interested, as well as by freely interpreting Steve's behaviour as conclusive proof that Steve loves him.

More generally, as Mele (2001) rightly describes, self-deceived subjects mistreat evidence in a number of ways. For example, self-deceived subjects may *positively misinterpret evidence* (interpret evidence that doesn't support *p* as evidence supporting *p*), they may *negatively misinterpret evidence* (interpret as not supporting *p* evidence that in fact supports *p*), they may *selectively focus on or attend to evidence* according to whether it supports *p* (selectively focus on evidence supporting their belief that p and/or fail to attend to evidence counting against *p*), or they may *selectively gather evidence* (overlook available evidence counting against *p* or actively search for less accessible evidence supporting *p*). According to Mele, these are all ways in which subjects' desire (or motivation) may lead them to misinterpret the evidence at hand, depending on the particular cases.

Motivationists, Mele included, generally assume that it is a *desire that p* that plays this biasing role. Nevertheless, there are variations and disagreements amongst motivationists regarding what kind of motivation can trigger this type of fallacious reasoning. Nelkin (2002), for example, argues that it is a *desire to believe p* rather than a desire that *p* that is at play in self-deception. Mele himself also admits, in his article on twisted self-deception (1999), that an emotion (e.g., jealousy) can play a similar biasing role that leads a subject to form a self-deceptive belief that that subject in fact wished were false.

How exactly these motivations, desires, and emotions might psychologically play a role in leading the subject to mistreat evidence is amply discussed in the literature (see Mele, 1997; 1999; 2001). In section III, we will focus on what Mele identifies as one of the underlying psychological mechanisms leading self-deceived subjects to mistreat evidence. But before turning to this aspect of self-deception, I will first present a different type of cases: high-stakes cases borrowed from the literature on pragmatic encroachment. As we will see, contrary to self-deceptive subjects, subjects presented in this type of case do not seem so irrational.

II. PRAGMATIC ENCROACHMENT

Contrary to what traditional views in epistemology assume, defenders of pragmatic encroachment argue that knowledge doesn't purely depend on truth-related factors but also varies according to the subject's practical interests—i.e., what is at stake for the subject in a given situation[6] (Stanley, 2005; Hawthorne, 2004; Fantl and McGrath, 2002; 2007; Hookway, 1990). In other words, advocates of the pragmatic-encroachment thesis defend the idea that a difference in the subject's practical circumstances, that a subject's interests and related risks, can *encroach* on the subject's epistemic state. This, of course, drastically departs

from the default view in epistemology, *purism*, according to which only truth-related factors such as truth, justification, reliability, and so on determine alone whether a subject is in a position to know.[7]

One way of motivating this original position is by appealing to a set of cases, the most famous of which are the so-called *bank cases*, originally presented by DeRose (1992).[8]

> *Low Stakes*. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. It is not important that they do so, as they have no impending bills. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Hannah remembers the bank being open on Saturday morning a few weeks ago, so she says, 'Fortunately, it will be open tomorrow, so we can just come back.' In fact, Hannah is right—the bank will be open on Saturday.
>
> *High Stakes*. Hannah and her wife Sarah are driving home on a Friday afternoon. They plan to stop at the bank on the way home to deposit their paychecks. Since their mortgage payment is due on Sunday, they have very little in their account, and they are on the brink of foreclosure, it is very important that they deposit their paychecks by Saturday. But as they drive past the bank, they notice that the lines inside are very long, as they often are on Friday afternoons. Hannah remembers the bank being open on Saturday morning a few weeks ago, so she says, 'Fortunately, it will be open tomorrow, so we can just come back.' In fact, Hannah is right—the bank will be open on Saturday. (Schroeder, 2012)

The reason why these cases support pragmatic encroachment is because our intuitions about whether the subject knows whether the bank will be open seem to shift from one case to the other. While we all tend to agree that Hannah knows that the bank will be open in the low-stakes case, many tend to disagree that Hannah is in a position to know that the bank will be open when the stakes are raised. Only, to admit this shift in intuition, to say that what is at stake for Hannah and her wife affects whether Hannah is in a position to know or not, amounts to denying purism and conceding that practical factors actually play a role in determining whether a subject is in a position to know.[9] For, if purism were true, Hannah would know that the bank will be open on Saturday morning both in the low-stakes case and in the high-stakes case, since the only variation between the cases is what is at stake for Hannah and her wife—i.e., the practical costs of being mistaken. Evidence, on the contrary, is stable across the cases.

This shifting intuition can, of course, be explained in a variety of ways. One might reject the validity of these intuitions and argue in accordance with purism that, if the subject knows *p* in one case, the subject must also know *p* in the other.

Another might argue that what is actually going on in the high stakes case isn't that the subject isn't in a position to know, or that the subject doesn't believe the target proposition, but rather that the subject isn't in a position to act rationally on the basis of this belief. There are indeed many ways of dealing with these cases, some of which are sympathetic to purism, others plainly rejecting it. As Engel (2009) notices, there are even different varieties of pragmatic encroachment, varying along two dimensions: (i) the kind of epistemic notion upon which these factors impede (whether it is knowledge, justification, belief, or rationality); and (ii) the degree of this influence.

Since my focus here concerns the rationality of beliefs, I will follow Schroeder's (2012) contribution, where he articulates pragmatic encroachment as a thesis about the rationality of beliefs. In his paper, Schroeder explores how we can make sense of the idea that practical considerations could encroach on knowledge. His solution to this puzzling idea is to say that practical considerations don't directly affect knowledge; rather, there is pragmatic encroachment on knowledge only because there is pragmatic encroachment on epistemic rationality. Given that epistemic rationality is a condition of knowledge, if high stakes defeat epistemic rationality, they defeat knowledge by defeating epistemic rationality. On his view, then, in the high-stakes case Hannah isn't in a position to know because she isn't in a position to rationally hold the belief that the bank will be open on Saturday.

As I said, pragmatic encroachment is a controversial position, and I do not wish to defend it here, since this would go far beyond the purposes of this paper. All I need to assume for the purposes of this paper is that there are at least some high-stakes cases in which the subjects would be more rational to suspend their belief, given the high stakes, than to believe *p*, and that, were they to believe *p*, their belief would be irrational. If you are not convinced that the bank cases meet these criteria, consider the following case.

> *Molly*
> Molly is about to pick some mushrooms in the forest to prepare the evening dinner for her family. She has good evidence that most if not all mushrooms in this area are edible (let us say her grandmother is a knowledgeable mushroom hunter who told her so). All the same, Molly is aware that many mushrooms are highly poisonous and that mistakes in identifying them can happen. Because Molly is motivated to avoid making a lethal mistake (she might die or poison her family as a result of this mistake), Molly suspends her belief and decides to check her *Mushroom Book* before serving dinner to find out whether the mushrooms she picked are edible.

In this case, there is less controversy about what exactly is going on, since it is explicitly mentioned that Molly suspends her belief. All the same, I think this is a plausible and ordinary case: sometimes, when the stakes are high, we suspend belief for the very reason that mistakenly holding the belief in question could

bear disastrous consequences. Following Schroeder's (2012) argument, as well as our intuitions, I think we can easily assume that there is something significantly more rational, if not plainly rational, in Molly's suspension of belief contrasted with Fernando's self-deceptive belief described in section I.[10]
Bearing this in mind, we will now see how comparing cases of self-deception with high-stakes cases of this type can generate a(n apparent) dilemma about the rationality of beliefs. I will thus begin by presenting the belief-formation mechanism that Mele (1997; 1999; 2001) appeals to, to explain how self-deceptive beliefs are formed. I will then argue that this very mechanism can also explain high-stakes cases.

III. THE (APPARENT) DILEMMA

As I suggested, both self-deception and high-stakes cases can be explained with reference to the same type of belief-formation mechanism, a type of belief-formation process under the influence of pragmatic considerations. This mechanism might not be sufficient for explaining all the puzzling aspects of self-deception, nor might it be capable of explaining all types of cases. All I claim is that it is sufficient to explain at least some cases, and that this is sufficient for generating a dilemma about the rationality of beliefs. What is more, it also seems to be this pragmatic influence that renders the resulting belief irrational. Only, if this is the case, then it becomes unclear why high-stakes cases aren't cases in which the subject is being irrational—a conclusion that I deem unlikely. Whereas self-deceptive beliefs are considered a paradigm of epistemic irrationality, high-stakes cases are (as we saw section II) at least somewhat more rational than cases of self-deception. Before articulating the dilemma, I will thus begin by explaining why I think both types of cases can be explained by the same type of mechanism.

In describing the psychological mechanisms underlying the various ways in which a subject may be led to misinterpret evidence in a motivationally biased way, Mele (1997; 1999; 2001) refers to what he labels the *Friedrich-Trope-Liberman (FTL) model* of lay hypothesis testing and ordinary reasoning, inspired by Friedrich's (1993) and Trope and Liberman's (1996) findings in psychology. The FTL model, alongside a variety of biases Mele mentions (1997; 1999; 2001), contributes to explaining the distorted ways in which self-deceived subjects mistreat evidence, just as Fernando does in this example.

According to the FTL model, subjects are pragmatic thinkers whose cognitions are first and foremost concerned with minimizing costly errors rather than driven by an interest for truth. This way of testing hypotheses and forming beliefs incorporates the subject's practical interests into the way in which the subject searches, gathers evidence, and forms or rejects beliefs. This mechanism for hypothesis testing and belief formation isn't purely truth oriented in the way scientific hypothesis testing would be. On the contrary, the type of hypothesis testing displayed in everyday reasoning is largely influenced by the practical costs thinkers associate with the possibility of forming a false belief or rejecting

a true one. The subjects' motivation for avoiding costly errors is implemented in their belief-formation process through a system of thresholds for belief formation and rejection that determine the amount of evidence required for forming or rejecting a belief. These two thresholds—which needn't be symmetrical—vary according to the costs of mistakenly believing (forming a false belief) or mistakenly rejecting a belief (rejecting a true belief), as well as according to the costs of information (that is, the costs of gathering further evidence, including resources, time, and effort) (Trope and Liberman, 1996; Mele, 2001). For example, a subject who associates high costs with falsely believing *p* will have a high threshold for forming this belief. If the costs of information are reasonable relative to the costs of falsely believing *p*, the subject will search for more evidence about this matter. The costs of falsely believing something aren't limited to material consequences. As Friedrich (1993) notes, self-serving motives, such as maintaining one's self-esteem, can also play a role in the determination of errors to be avoided. Thus, the practical costs of falsely believing include a wide range of costs, including psychological ones (those linked to the subject's self-image, for instance) as well as material ones (such as poisoning).[11]

To see how this model functions, consider the following example. Imagine you are walking in your back garden and notice a few rhubarb leaves growing there. You know that this plant is somehow edible, but you aren't quite sure which part of the plant you can actually eat, or whether some of it might be indigestible. You also remember eating delicious rhubarb pie but cannot quite remember which part of the plant had been used in the preparation. You form the following hypothesis.

(H) *Rhubarb leaves are edible.*

Knowing that the risks of forming a belief on the basis of this hypothesis might be fatal (you are aware that some plants are poisonous), you are cautious and do not jump to conclusions. Instead, you test the hypothesis by searching for disconfirming evidence—that is, evidence that rhubarb leaves are poisonous—for this will give you the best chances of avoiding forming a false belief. And because the costs of falsely believing that rhubarb leaves are edible are high (i.e., the risk of suffering some sort of food poisoning), the amount of evidence you gather before forming the belief in question is important: you do not want to make a mistake about this particular matter and poison yourself or your guests. In other words, the higher the stakes or the costs of falsely believing *p*, the more evidence one needs in order to form a given belief.

The FTL model, as Mele points out, participates in the formation of self-deceptive beliefs. If we apply this model to a case of self-deception such as the one presented above, we can see how the model predicts and explains the distorted ways in which some self-deceived subject forms his or her belief. In this particular case, Fernando is self-deceived in believing that Steve is in love with him. Fernando not forming the appropriate belief on the basis of the evidence can be explained by the fact that he associates high costs with forming the belief that

Steve isn't romantically interested, and low costs with falsely believing that his love is requited. The truth (not-*p*) might break Fernando's heart and lead to severe emotional consequences for him, whereas the costs of falsely believing that his love is requited are low, as the belief serves only to maintain him in a happy bubble.[12] These costs of believing not-*p* (i.e., that Steve isn't in love with him) thus set a high threshold for forming the appropriate belief, and the amount of evidence required for believing not-*p* is correspondingly high. If this is correct, then at least some cases of self-deception, if not all, can be explained by referring to this type of belief mechanism as the one described by the FTL model.[13]

Self-deceptive beliefs, we have seen, are formed through a motivationally biased process. The FTL model presented above explains how a subject can be led to mistreat evidence in order to avoid forming costly false beliefs.[14] This mechanism for belief formation participates in making the belief irrational since it causes the subject to not form the appropriate belief. The motivation to avoid forming costly false beliefs—as well as, more generally speaking, the influence of practical considerations on beliefs—is typically deemed irrational. This suggests that the FTL model can at least explain some cases of self-deception: the costs of falsely believing function as a motivation to avoid certain undesired consequences that influences the subject's treatment of evidence and leads the subject to form a self-deceptive, irrational, belief.

Interestingly, these high-stakes cases can easily be explained by referring to the FTL model, with the very same mechanism that was responsible for the generation of irrational beliefs in cases of self-deception. In high-stakes cases such as the one described above, the costs associated with falsely believing *p* are high: Hannah and Sarah will be in a very uncomfortable situation if they fail to deposit their paycheques before Sunday. Since these costs are high, the threshold for belief formation is equally high and the amount of evidence required for believing that the bank will be open on Saturday is greater than that in the low-stakes case. This can explain why the amount of evidence can be sufficient for belief formation in the low-stakes case, without being sufficient when the stakes are high. Hannah's concern for avoiding costly false beliefs influences her threshold for belief formation and leads her to suspend her belief until further evidence is gathered. The same goes for Molly: her position is perfectly analogous to the example used to illustrate the FTL model. It is because the costs associated with falsely believing that the mushrooms are edible are extremely high that the threshold for belief formation is equally high. This results in Molly not forming the belief that the mushrooms are edible until she gathers further evidence from her *Mushroom Book*.

But, contrary to cases of self-deception such as the one described previously, the stakes here influence Hannah and Molly's beliefs in a rational way. Indeed, as I said, cases of this sort are used in the context of debates on pragmatic encroachment to bring about the intuition that there's a good reason for subjects to suspend their belief and search for further evidence. In what follows, I rely on

the intuition that in such cases the perceived costs of falsely believing actually have a positive effect on the subject's belief, in the sense that, even if we understand high-stakes cases by referring to the system of adaptive thresholds described by the FTL model, this does not affect the subject's rationality.
Cases of self-deception thus provide us with an example in which pragmatic factors (in particular, the costs of forming a false belief) influence the subject's belief formation. This influence is what makes the belief irrational and insensitive to evidence. Take the example of Fernando again. Fernando believes (falsely) that Steve is in love with him. This belief results from the fact that Fernando associates high costs with falsely believing that Steve doesn't love him and associates low costs with believing that Steve does love him. According to the standard understanding of self-deception, what makes self-deceptive beliefs irrational lies precisely in the fact that the beliefs are formed through a motivationally biased process: a process under the influence of practical factors (such as desires, the motivation to avoid costly false beliefs, and so on), an influence that leads the subject to mistreat the evidence at hand. But, as we have seen with cases of pragmatic encroachment, the same mechanism can influence a subject's belief without affecting that subject's rationality. On the contrary, this influence seems warranted: when the stakes are high, it is more rational for subjects to suspend their belief until further evidence is gathered, rather than holding onto it. Indeed, were Hannah or Molly to believe *p* in the high-stakes situation—that is, were the stakes to not influence her belief—her belief would be irrational. In these cases, it might be more rational for Hannah or Molly to suspend her belief. Both sets of cases are cases in which the practical costs of forming false beliefs affect the subject's belief through what seems to be a system of thresholds for belief formation and rejection. But, while in the first case this influence seems to make the subject's belief irrational,[15] in the second this influence does not undermine the subject's rationality. The puzzling aspect of this comparison is precisely that this biased mechanism is what determines the rationality of the belief in the first case. If we accept these claims, we are then led to the following dilemma about rationality.

> *Dilemma*
> *Either* this mechanism produces irrational beliefs, and both self-deception and cases of pragmatic encroachment are irrational
> *or* this mechanism affects beliefs in a rational way, and both cases of pragmatic encroachment and cases of self-deception are rational.

There are of course three options for responding to this dilemma. We can decide either to embrace either the first horn or the second or to reject the dilemma by explaining why it doesn't hold. In what follows, I assume that neither horn is likely (that self-deception is indisputably irrational and that there is something rational—or at least significantly more rational—about suspending one's belief in high-stakes cases). If we want to find a way out of this dilemma, we need to deny that the rationality of one's belief has anything to do with being formed via the mechanism of adaptive thresholds described by the FTL model. If this line of thought is correct, the dilemma is merely apparent, and the irrationality of

self-deception lies elsewhere: it has to be found in a feature of self-deception that cases of pragmatic encroachment do not possess.

IV. HYPOTHESES AND SOLUTIONS

In this final section I spell out several differences between the two cases to find out which element, present in self-deception but not in cases of pragmatic encroachment, is responsible for the irrationality of self-deceptive beliefs. The hypotheses listed in the table below represent some, and hopefully all, of the relevant differences between the two cases. I do not intend to say that these hypotheses are mutually exclusive or that they cannot be combined to furnish a full explanation of this difference in rationality, nor do I mean to say that they bear no logical connection to one another. As we will see, I think some of the hypotheses are in fact related.

Hypothesis	*Self-deception*	*Pragmatic Encroachment*
H1	Falsity	Truth
H2	Formation	Suspension
H3	Quick reasoning	Slow reasoning
H4	Costs of believing *p*	Costs of *p* being false

Hypothesis 1: Truth versus falsity

According to the first hypothesis, the fact that in cases of self-deception the subject's belief is false, whereas in cases of pragmatic encroachment the belief is true, can explain why cases of self-deception are irrational and high stakes cases are not. This difference might seem like a very trivial and simplistic one, for, as we know, falsity of the belief isn't a necessary condition for self-deception. It is indeed easy to see how, even though in most cases the self-deceptive belief will happen to be false, it could also accidentally be true. Irrationality and falsity are independent from one another, and a subject could be warranted to believe *p* even though *p* is false.

Hypothesis 2: Suspension versus formation

One can draw attention to the sort of impact these practical costs have on the subject's belief. In one case, self-deception, the subject's belief results from practical considerations, whereas in the case of pragmatic encroachment the subject merely suspends his or her belief. In his (2012) article, Schroeder argues that reasons to withhold belief are necessarily nonevidential. He writes:

> Why is it that reasons to withhold cannot be evidence? It is because the evidence is exhausted by evidence which supports p and evidence which

> supports ~p. Consequently, the reasons to withhold must come from somewhere else. So they cannot be evidence. (Schroeder, 2012)

If Schroeder is correct, and if reasons to believe must be evidential, whereas reasons to withhold belief can only be practical, then this might explain the difference between the two sets of cases. This argument would allow us to explain why the influence of practical costs makes beliefs irrational in the case of self-deception, but not in typical cases of pragmatic encroachment, if, as I said, cases of self-deception are cases in which subjects *form* beliefs, and cases of pragmatic encroachment are cases in which subjects *suspend* beliefs. But is this true?

Although I think all cases of the second type—that is, cases of pragmatic encroachment—are cases in which subjects suspend their belief, the question of whether there can be cases of self-deception in which the self-deceptive subjects also suspend their belief is less straightforward. Indeed, despite self-deception being mostly defined as a process by which a belief is *formed,* applying only to cases in which a subject is self-deceived in believing *p*, this definition might be too restrictive (cf. the definition in section I). It does not seem to be true that one cannot be self-deceived in simply refusing to believe something, in suspending one's belief in the face of strong evidence instead of forming the warranted belief. There are cases in which the subject suspends belief for practical reasons, but these cases fail to be rational, and these cases are cases of self-deception. Here is an example of this sort.

> *Suspension Jo*
> Jo is meeting Laurie for a drink by the lakeside. She has very good reasons to believe that this is a date and that Laurie is romantically invested: he regularly sends her handwritten letters and red roses and bakes her delicious cookies. Their best friend also confirmed that Laurie does this only when he's madly in love. Nevertheless, Jo refuses to believe that Laurie is romantically interested, and suspends belief. Her reason for doing so is that the mere thought of him liking her in return is overwhelming and puts her in a general state of emotional distress.

Although the case doesn't fit with the definition of self-deception presented above (insofar as the definition mentions only cases in which the subject forms or holds a belief), I think some might still hold the intuition that Jo is self-deceiving in suspending belief. Imagine a discussion between Jo and her sister Beth: Beth is insisting that Laurie is in love with Jo, but Jo refuses to see the evidence supporting this belief. Would Beth be tempted to tell her sister that she is self-deceived? Maybe. Maybe not. Some might want to argue that self-deception necessarily involves forming a belief and that merely suspending one's belief, however irrational this may be, cannot constitute a genuine case of self-deception. However, it isn't crucial that we agree on calling this a case of self-deception; all we need is to agree that Jo is being irrational in refusing to believe that Laurie is romantically interested in her, which I think she is. If so, then the third

hypothesis does not account for the intuitive difference in rationality between the two sets of cases.

Hypothesis 3: Quick versus slow[16]

Another way of explaining the difference in rationality between the two cases is by referring to the way in which the beliefs are formed—i.e., the type of reasoning by which they are formed. One could argue that for a belief to be self-deceptive it should not be sufficient that the belief-formation process be influenced by the costs of falsely believing. In addition to this, the belief should be formed through a certain type of reasoning. Accordingly, the threshold variation isn't what causes beliefs to be irrational in cases of self-deception; rather, it is the biased treatment of the evidence that causes the belief to be irrational. Mele (1997; 1999; 2001) refers to several cognitive biases that can be triggered by desires or emotions and that might participate in the formation of self-deceptive beliefs. Such biases include the confirmation bias, the vividness-of-information bias (Mele, 1997; 1999; 2001), and even the availability-heuristics bias (Mele, 1997; 1999; 2001).

By contrast, subjects in cases of pragmatic encroachment do not go through this type of process, and this is why their attitude isn't irrational. A more precise formulation of this idea can be given by referring to dual-process theory (Evans, 2010; Evans and Frankish, 2009; Frankish, 2010). Mainly developed by Tversky and Kahneman (1974), dual-process theory provides a schematic vision of our cognition divided into two systems, two distinct ways of reasoning and treating information: system 1 and system 2. The first system is intuitive: it is quick, implicit, and effortless. The second process on the contrary is reflective; it is slower and relies on controlled, analytic thinking. Cognitive biases are understood as belonging to system 1: they allow us to treat information effectively while investing little cognitive effort, but can lead to errors. System 2 is more rational, but requires more cognitive effort and time.

This second hypothesis draws upon the dual-process theory to explain what might be irrational about self-deception and rational in cases of pragmatic encroachment. The relevant difference between these cases has nothing to do with motivation. In fact, the difference has to do only with the way in which this motivation affects our thinking. In cases of self-deception, what makes the belief irrational is that it results from a certain type of reasoning process, an intuitive, associative way of treating information, that results in a false belief. It is the motivation, the emotion, or the desire that leads the subject to think in system 1: the self-deceptive subjects don't reason analytically. And this is what differentiates them from the subjects in cases of pragmatic encroachment. Subjects in cases of pragmatic encroachment engage in slow, analytical thinking. Their motivation for reevaluating their belief might be a practical one, but their thinking isn't biased by their motivation. It isn't the belief in itself that is irrational, but only the process by which it is formed. According to this hypothesis, the influence of the practical costs of falsely believing in one's thresholds for belief

acceptance and rejection doesn't play a role in determining whether the resulting belief is rational or not: what matters is the type of reasoning or thinking that leads to this belief.
This hypothesis fails to explain the difference for the following reason: many rational beliefs are formed through system 1; therefore, relying on biases and heuristics does not necessarily make a subject's belief irrational. If this is correct, as I think it is, then why would this type of reasoning be what determines the belief's rationality only in this specific case? Since the answer to this question forces us to search for a further criterion, it cannot be a convincing solution to the dilemma.

Hypothesis 4: Costs of believing p versus costs of p being false

According to the fourth hypothesis, the relevant difference between the cases concerns the kind of practical costs at play. Indeed, cases of self-deception are cases in which the costs influencing the subject's threshold for belief are those related to holding the belief itself rather than the costs related to falsely believing *p*. In other words, what influences the subject's belief doesn't depend on the falsity of the belief; it merely depends on holding the belief regardless of whether it might be true or false. What matters to the self-deceived subject is to *believe p* no matter what the evidence suggests, regardless of whether *p* might be true or not. In the high-stakes cases, on the contrary, Hannah is concerned with the truth of *p*—what she wants to find out is whether *p* is true or not. Although the amount of effort invested in the inquiry is influenced by pragmatic considerations (the practical costs of falsely believing *p*), her interest is to reduce the possibility of making a costly error—what matters to her is that her belief be true.

This distinction in fact neatly overlaps one offered by Jordan (1996), who presents two types of pragmatic arguments for belief formation: *truth-dependent* and *truth-independent* arguments. Truth-dependent arguments are pragmatic arguments for believing something because, if *p* happens to be true, then the practical benefits of believing *p* will be great. They are truth dependent precisely because these practical benefits depend on *p* being true. If *p* turned out to be false, then there would be no benefits to holding the belief in question. Truth-independent arguments, on the contrary, are pragmatic arguments for believing *p*, the benefits of which do not depend on p being true: the benefits gained by believing p hold whether or not *p* turns out to be true (Jordan, 1996). If you think of Fernando again, you will realize that what influences his threshold for belief formation doesn't depend on what will happen were he to believe falsely, but depends on the emotional costs of believing that Steve doesn't love him, something he is afraid to admit. The self-deceived subject isn't interested in finding out whether *p* is true or not—this is precisely not the point of their inquiry.

A wining hypothesis?

What conclusions can we draw from the evaluation of these hypotheses? Hypothesis 1 isn't convincing as it merely indicates an accidental feature of self-

deception. Hypothesis 2 does set up an interesting basis for explaining the difference between the cases, but assumes that there are no cases of self-deception in which the subject suspends belief, which is misleading. It is difficult to see how hypothesis 3 could ground a proper difference in rationality since many rational beliefs are formed through automatic reasoning. But as we will see, maybe hypotheses 2 and 3 haven't said their last word. Finally, hypothesis 4, I think, has more potential. First, it smoothly applies to both cases: In the case of self-deception, the subject's belief is influenced by the costs of believing p (regardless of whether p is true or not) rather than by the costs of falsely believing p. In the high-stakes case, on the contrary, Hannah's and Molly's thresholds for belief aren't influenced by the costs of believing but by the costs of falsely believing p. But this doesn't seem to be the full story.

Further support for this idea can be found in found in Kunda's (1990) and Kruglanski's (Kruglanski, 1980; Kruglanski and Ajzen, 1983; Kruglanski and Klar, 1987) works on motivated reasoning. In her famous paper, Kunda also distinguishes between two types of motivations: the *motivation to arrive at an accurate conclusion* (whatever the conclusion may be) *versus the motivation to arrive at a particular, directional, conclusion*.[17] While the first "enhances use of those beliefs and strategies that are considered most appropriate," the second "enhances use of those that are considered most likely to yield the desired conclusion" (Kunda, 1990). If Kunda is correct in this, then the type of motivation also influences the type of reasoning, not exactly in the sense described under hypothesis 2, but in the following way: directionally motivated subjects will tend to rely on ways of reasoning that allow them to "construct seemingly reasonable justifications for these conclusions" (Kunda, 1990). Subjects whose reasoning is directionally motivated do not only tend to rely on biased reasoning, but they also seem to "pick and choose" reasoning strategies likely to lead them to form the desired belief.

This points to another interesting difference between self-deceptive subjects and high-stakes subjects. In cases of self-deception, subjects do not recognize their motivations as being part of the reason why they come to believe *p*. On the contrary, they often seem unaware of this causal connection. Self-deceived people typically take the evidence to support their false belief, whereas it seems part of high-stakes subjects' position to be aware of the fact that they are suspending belief for practical reasons.

Finally, and given what I have just said, although I did argue that there were irrational cases of suspension of belief, it might well be the case that practical reasons bear a rational influence on beliefs only in cases of suspension.[18] In other words, this would mean that, although it could be warranted to suspend one's belief about whether *p* for practical reasons, one could never rationally believe *p* for practical reasons (no matter what type of reasoning one is using). In fact, this is compatible with what Schroeder (2012) suggests when he writes that reasons to withhold can only be nonevidential. The final question would thus be the following: Would we deem it rational for a subject to suspend belief for

truth-independent reasons? I suspect not. In the light of the points presented above, we could thus add the following: it might be rational to whithhold belief for practical reasons, if these reasons are truth-dependent in the sense specified by Jordan (1996).[19]

These comments might, of course, merely constitute the sketch of an answer. They might even, as a matter of fact, raise more questions than they dare to answer. All the same, I hope to have shown that the suggested dilemma doesn't really hold, that, despite these similarities, self-deception and high-stakes cases differ in significant ways.

That being said, let us finally make note of a few implications for theories of self-deception and pragmatic encroachment. First, we need to recognize that, although Mele might be correct in recognizing the FTL model as a mechanism by which we can explain self-deception, it seems misleading to think that any type of motivation might play this role. Indeed, the motivation at play in cases of self-deception seems more likely to be a motivation to avoid forming beliefs *tout court*, rather than a motivation to avoid forming costly *false* beliefs. This might be a reason to prefer a motivationist account, such as Nelkin's (2002), that argues that self-deception should be defined as a desire to believe *p* rather than a desire that *p,* as most motivationists put it, or an account such as the one suggested by Lauria et al. (2016), according to which self-deception is best understood as a type of affective coping. These accounts, as well as the argument laid out above, seem to suggest that the motivation inherent to self-deception is a motivation linked to the costs of believing rather than related to the truth of *p*, something that isn't obvious in Mele's account.

Second, this should also have implications for our conception of what makes self-deceptive beliefs irrational. Indeed, as we have seen, the FTL model in itself isn't sufficient for explaining why we deem self-deceptive beliefs to be irrational since the same mechanism also leads to rational cases. As I suggested, it is important to distinguish between the costs of believing and the costs of believing falsely, and thereby acting as if *p*. Although I here suggested that this difference can explain why self-deception is irrational, whereas cases of pragmatic encroachment are not, there is more to say about how and why, precisely, these two types of influences ground rationality.

Third, on the pragmatic encroachment side of the discussion, this proves that the cases used to support the thesis lack detail. Although we do seem to have the intuition that these subjects aren't self-deceived, the lack of precision about what type of practical considerations and reasoning might be compatible with this rationality allows us to question the belief processes involved. In the absence of detail, one might just as well use this as an argument against pragmatic encroachment in order to show why these cases are perfectly analogous to self-deception.

Last but not least, I think this discussion has established that speaking in terms of the influence of practical factors alone isn't sufficient for explaining why a

belief might be irrational. Without heading towards the pragmatists' camp and saying that it is rational to believe whatever maximizes utility, for example (see Rinard, 2015; 2017 for a discussion and defence of pragmatism), I claim that it seems insufficient, even within a more evidentialist framework, to simply posit that only evidential considerations play a role in determining the rationality of one's beliefs.

CONCLUSION

I have argued that cases of self-deception and cases of pragmatic encroachment can be explained by reference to the same mechanism—namely, the FTL model for belief formation and lay hypothesis testing. This mechanism shows that a subject's motivation for avoiding costly false beliefs not only explains the way in which the belief is formed, but also seems to explain why this belief is irrational—insofar as the motivation to avoid costly false beliefs thereby leads the subject to mistreat the evidence at hand. On this basis, I argued that by accepting this we are forced into a dilemma about the rationality of beliefs according to which either self-deception is rational or cases of pragmatic encroachment are irrational. Finally, I presented several hypotheses about how to solve this dilemma and argued that the type of motivation at play holds a central role in distinguishing the two types of cases.

ACKNOWLEDGEMENTS

I am grateful to Natalie Ashton, Jie Gao, Marie van Loon, Alfred Mele, Anne Meylan and Jennifer Nagel, as well as the two anonymous referees at *Les ateliers de l'éthique*/*The Ethics Forum* for their insightful comments. Earlier versions of this article were presented at "Self-Deception: What It Is and What It's Worth," University of Basel; "Ateliers du GRE," Collège de France; the "Epistemic and Practical Rationality" workshop, University of Fribourg; and the "European Epistemology Network" conference, Vrije Universiteit, Amsterdam. I thank the audiences at these venues for the stimulating discussions that followed these presentations. Research for this article was supported by the Swiss National Science Foundation (SNSF) Professorship grant "Irrationality" #PP00P1_157436 and the Doc.Mobility grant "Believing Under the Influence" #P1BSP1_181672.

NOTES

[1] To my knowledge, Gao (n.d.) is the only author who points out and discusses this similarity.

[2] By "high-stakes cases" I mean to say cases described and used in the literature on pragmatic encroachment. The cases are often used as a motivation for rejecting purism, the idea that knowledge or other epistemic states are purely truth related (see section II for more detail).

[3] This second condition helps mainly to distinguish self-deception from wishful thinking, often defined as cases in which the subject is only unwarranted in believing p (Szabados, 1973; Van Leeuwen, 2007).

[4] In one of his footnotes (2006, p. 115), Mele writes "the requirement that p be false is purely semantic. By definition, one is *deceived in* believing that p only if p is false, the same is true of being *self-deceived in* believing that p. The requirement does not imply that p's being false has special importance for the dynamics of self-deception. Biased treatment of data may sometimes result in someone's believing an improbable proposition, p, that happens to be true" (see also Mele, 1987, p. 127-128).

[5] Van Leeuwen (2007) describes "not-p" as the "doxastic alternative". Attitudes towards the doxastic alternative vary from one theory of self-deception to another. On Mele's view, it is not necessary that the subject *believes* the doxastic alternative to be self-deceived. However, condition (b) seems to describe the self-deceived subject as somehow possessing evidence supporting the doxastic alternative.

[6] Most of versions of pragmatic encroachment primarily focus on "stakes." Nonetheless, some philosophers (see Anderson, 2015; and Gerken, 2011) argue that other types of factors such as urgency, the availability of alternative evidence, social rules, and conventions can play a similar role.

[7] Pragmatic encroachment on knowledge, for example, must be understood as a thesis about the metaphysics of knowledge rather than as a thesis about the pragmatics of the verb "to know."

[8] For similar cases, see Cohen, 1999; Fantl and McGrath, 2002; and McGrath, 2018.

[9] The traditional position is usually referred to as purism. Purism can be spelled out as follows: "For any two possible subjects S and S', if S and S' are alike with respect to the strength of their epistemic position regarding a true proposition p, then S and S' are alike with respect to being in a position to know that p" (Fantl and McGrath, 2007).

[10] One could, of course, reject the assumption I am establishing: that is, that there are no such cases of subjects suspending belief in high-stakes situations because believing *p* in these circumstances would be irrational. If this were the case, there would be no dilemma concerning the rationality of beliefs, for the rationality of the high-stakes subjects wouldn't concern their beliefs, but their action. Against this objection, one could invoke the tight link between

action and belief, or between action and knowledge. For example, functionalists about beliefs may argue that to believe *p* is to be disposed to act as if *p* (see Ganson, 2007) for a discussion of the double function of beliefs in relation to this issue. Defenders of pragmatic encroachment often invoke something they call the "knowledge-action principle." Roughly put, this principle states that if *S* knows *p* then *S* is in a position to act as if *p* (see Fantl and McGrath, 2002; Williamson, 2005; Stanley and Hawthorne, 2008, for different formulations of this idea). This principle is used to reinforce the idea that if *S* isn't in a position to act when the stakes are high, then *S* doesn't know that *p*.

[11] A similar idea can be found in James's (1897) work, in which he mentions "two duties in inquiry": (i) avoiding false beliefs and (ii) forming true beliefs. Depending on which of these two duties the subject takes to be his or her primary concern, the subject will alter his or her inquiry and treatment of the evidence relative to whether p. If the subject is primarily concerned with (i) avoiding falsehood, the amount of evidence required for believing p will be greater, whereas if the subject's primary concern is (ii) forming true beliefs—and relieving himself or herself from a state of agnosticism—that subject will then form a belief on weaker grounds. More recently, Nagel introduced two psychological elements that influence belief formation and epistemic inquiry: epistemic anxiety (2010) and need for closure (2008). These two "forces" vary in function of the practical interests of the subject in a given situation (Nagel, 2008). Epistemic anxiety is the emotive response resulting from the perceived costs in being mistaken about a particular matter; the response consists in the subjects regulating their cognitive effort and adapting their cognitive strategy by relying on more deliberate and controlled cognition (Nagel, 2008; cf. Tversky and Kahneman, 1974). One's level of "need for closure" on the other hand (cf. Kruglanski and Webster, 1996) refers to the threshold of belief (or desired "levels of confidence"; Nagel, 2010) at which a subject settles and forms a given belief.

[12] There might be costs other than the ones mentioned here. These costs could range from further subjective, psychological costs to more objective costs. For example, one could consider the costs of causing distress to Steve by misinterpreting his behaviour. I think the issue of determining how to narrow down the relevant costs hasn't yet been completely clarified: do these costs depend only on the subject's interests and primary concerns? For purposes of simplicity, let's here assume that the relevant costs are the ones described above.

[13] This mechanism described as the FTL model in Mele's words is in fact very close to the general functioning of adaptive cognition. Roughly speaking, adaptive cognition is the idea that our cognition, the way in which we treat information, test hypothesis, and form beliefs is cognitively adaptive in the following sense: "agents adapt their cognitive efforts" (and resources) "to how they represent the practical factors relevant to the task at hand" (Gerken, 2017). This means that the ways in which agents perceive their practical situation influence their cognition in a significant way. Although there is significant disagreement, as Gerken (2017) notices, "there is a wide agreement that our metacognitive procedures adapt the cognitive resources that we deploy for a given task to how they represent the practical factors associated with it." The two central aspects of adaptive cognition are the following: (i) how much cognitive effort one is willing to allocate to a given cognitive task as well as (ii) how much evidence one needs in order to form or reject a given belief, both very according to one's practical stuation.

[14] I thank the reviewers for noting that twisted self-deception may present a challenge to any account heavily relying on this type of mechanism. However, many accounts of self-deception face this challenge, and it might be the case that Mele takes FTL to sometimes, but not necessarily, play a role in the formation of self-deceptive beliefs. And it is sufficient for our argument that FTL sometimes produces irrational beliefs.

[15] Although Mele does rely on the FTL model to explain how self-deceptive beliefs come about, it might be the case that neither Trope nor Friedrich nor Liberman would agree with the idea that, according to the FTL model (which, I recall, is a generalization provided by Mele, 1997; 1999; 2001), the motivation to avoid costly false beliefs may result in irrational beliefs. This

might be true, for example, for anyone working with a slightly different notion of rationality (i.e., an evolutionary notion, for example) than the one assumed throughout this discussion.

[16] I am grateful to Alfred Mele for suggesting this third hypothesis.

[17] For more work on this distinction see Kruglanski, 1980; Kruglanski and Ajzen, 1983; Kruglanski and Klar, 1987; see also Chaiken, Liberman, and Eagly, 1989; Pyszczynski and Greenberg, 1987.

[18] I here set aside the pragmatist (or nonevidentialist) idea that it can be rational to believe for nonevidential reasons because accepting the truth of pragmatism is incompatible with my beginning assumption that self-deception is irrational. Indeed, if we accept that it can be rational to believe p because believing p leads to positive consequences, for example (cf. Rinard, 2015; 2017), it then becomes unclear why we would still consider most cases of self-deception irrational.

[19] It could be interesting to think about Pascal's Wager here. Pascal's wagerer decides to believe in God because he or she thinks believing in God will lead to positive consequences whatever the truth is, whereas disbelieving (whether this means believing that God doesn't exist or suspending belief) will either lead to no consequences or lead to negative ones. Overall, this could be understood as an FTL belief. However, I do not think this would qualify as a self-deceptive belief for the following reasons: First, being the product of FTL isn't sufficient for qualifying as a self-deceptive belief (cf. conditions given in section 1). It isn't obvious—at least not to me—that Pascal's wagerer should in fact believe not-*p* (that God doesn't exist) in the sense that he or she has been presented with sufficient evidence for being warranted in believing not-*p*. Second, these final considerations about the difference between self-deception and high-stakes cases also shed light upon the kind of reasoning underlying self-deception, and it does not seem to me that Pascal's wagerer is similar in this respect. One might still want to argue that forming a belief for pragmatic reasons of this sort—regardless of whether one has evidence supporting this belief—is irrational, even though it might not be self-deceptive.

REFERENCES

Anderson, Charity, "On the Intimate Relationship of Knowledge and Action," *Episteme*, vol. 12, no. 3, 2015, p. 343-353.

Barnes, Annette, *Seeing through Self-Deception*, New-York, Cambridge University Press, 1997.

Chaiken, Shelly, Akiva Liberman, and Alice H. Eagly, "Heuristics and Systematic Information Processing within and beyond the Persuasion Context," in Uleman, J. S. and J. A. Bargh (eds.), *Unintended Thought: Limits of Awareness, Intention, and Control*, New York, Guilford Press, 1989, p. 212-252.

Cohen, Stewart, "Contextualism, Skepticism, and The Structure of Reasons," *Philosophical Perspectives*, vol. 13, 1999, p. 57-89.

Davidson, Donald, "Deception and Division," in LePore, E. and B. McLaughlin (eds.), *Actions and Events*, New York, Basil Blackwell, 1985.

———, "Paradoxes of Irrationality," Oxford, Clarendon Press, 2004.

DeRose, Keith, "Contextualism and Knowledge Attributions," *Philosophy and Phenomenological Research*, vol. 52, no. 4, 1992, p. 913-929.

Engel, Pascal, "Pragmatic Encroachment and Epistemic Value," in Haddock, A., A. Millar, and D. Pritchard (eds.) *Epistemic Value*, Oxford, Oxford University Press, 2009.

Evans, Jonathan. St. B. T., *Thinking Twice: Two Minds in One Brain*, Oxford, Oxford University Press, 2010.

Evans, Jonathan St. B. T., and Keith Frankish, *In Two Minds: Dual Processes and Beyond*, Oxford University Press, Oxford, 2009.

Fantl, Jeremy, and Matthew McGrath, "Evidence, Pragmatics, and Justification," *Philosophical Review*, vol. 11, no. 1, 2002, p. 67-94.

———, "On Pragmatic Encroachment in Epistemology," *Philosophy and Phenomenological Research*, vol. 75, no. 3, 2007, p. 558-589.

Frankish, Keith, "Dual-Process and Dual Theories of Reasoning," *Philosophy Compass*, vol. 5, no. 10, 2010, p. 914-926.

Friedrich, James, "Primary Error and Detection and Minimization (PEDMIN) Strategies in Social Cognition: A Reinterpretation of Confirmation Bias Phenomena," *Psychological Review*, vol. 100, no. 2, 1993, p. 298-319.

Funkhouser, Eric, "Do the Self-Deceived Get What They Want?" *Pacific Philosophical Quarterly*, vol. 86, no. 3, 2005, p. 295-312.

Gao, Jie, "Self-Deception and Pragmatic Encroachment, a Dilemma for Epistemic Rationality," manuscript, n.d.

Ganson, Dorit, "Evidentialism and Pragmatic Constraint on Outright Belief," *Philosophical Studies,* vol. 139, no. 3, 2008, p. 441-458.

Gerken, Mikkel, “Warrant and Action,” *Synthese*, vol. 178, no. 3, 2011, p. 529-547.

Hawthorne, John, *Knowledge and Lotteries*, Oxford, Oxford University Press, 2004.

Hookway, Christopher, *Scepticism*, London, Routledge, 1990.

James, William, The Will To Believe, and Other Essays in Popular Philosophy, New York, Longmans, 1897.

Johnston, Mark, “Self-Deception and the Nature of the Mind,” McLaughlin, B. and A. Rorty (eds.), 1988, p. 63-91.

Jordan, Jeff, “Pragmatic Arguments and Belief,” *American Philosophical Quarterly*, vol. 33, no. 4, 1996, p. 409–420.

Kahneman, Daniel, and Amos Tversky, “Judgement Under Uncertainty: Heuristics and Biases,” *Science*, vol. 185, no. 4157, 1974, p. 1124-1131.

Kunda, Ziva, “The Case for Motivated Reasoning,” *Psychological Bulletin*, vol. 108, no. 3, 1990, p. 480-498.

Kruglanski, Arie W., “Lay Epistemology Process and Contents,” *Psychological Review*, vol. 87, p. 70-87, 1980.

Kruglanski, Arie W., and Icek Ajzen, “Bias and Error in Human Judgment,” *European Journal of Social Psychology*, vol. 13, no. 1, 1983, p. 1-44.

Kruglanski, Arie W., and Yechiel Klar, “A View from the Bridge: Synthesizing the Consistency and Attribution Paradigms from a Lay Epistemic Perspective,” *European Journal of Social Psychology*, vol. 17, no. 2, 1987, p. 211-241.

Lazar, Ariela, “Deceiving Oneself or Self-Deceived? On the Formation of Beliefs Under the Influence,” *Mind*, vol. 108, no. 430, 1999, p. 265-290.

Lauria, Federico, Delphine Preissmann, and Fabrice Clément, “Self-Deception as Affective Coping: An Empirical Perspective to Philosophical Issues,” *Consciousness and Cognition*, vol. 41, 2016, p. 119-134.

McGrath, Matthew, “Defeating Pragmatic Encroachment,” *Synthese*, vol. 195, no. 7, 2018, p. 3051-3064.

Mele, Alfred, “Real Self-Deception,” *Behavioral and Brain Sciences*, vol. 20, no. 1, 1997, p. 91-136.

———, “Twisted Self-Deception,” *Philosophical Psychology*, vol. 12, no. 2, 1999, p. 117-137.

———, *Self-Deception Unmasked*, Princeton, Princeton University Press, 2001.

———, “Self-Deception and Delusions,” *European Journal of Analytic Philosophy*, vol. 2, no. 1, 2006, p.109-124.

Nagel, Jennifer, “Knowledge Ascriptions and the Psychological Consequences of Changing Stakes,” *Australasian Journal of Philosophy*, vol. 86, no. 2, 2008, p. 279-294.

———, “Epistemic Anxiety and Adaptive Invariantism,” *Philosophical Perspectives*, vol. 24, no. 1, 2010, p. 407-435.

Nelkin, Dana K., “Self-Deception, Motivation, and the Desire to Believe,” *Pacific Philosophical Quarterly*, vol. 83, no. 4, 2002, p.384-406.

Pears, David, *Motivated Irrationality*, Oxford, Clarendon Press, 1984.

Pyszczynski, Tom, and Jeff Greenberg, “Toward an Integration of Cognitive and Motivational Perspectives on Social Inference: A Biased Hypothesis-Testing Model,” in Berkowitz, L. (ed.), *Advances in Experimental Social Psychology*, New York, Academic Press, vol. 20, 1987, p. 297-340.

Rinard, Susanna, “Against the New Evidentialists,” *Philosophical Issues*, vol. 25, no. 1, 2015, p. 208-223.

Rinard, Susanna, “No Exception for Belief,” *Philosophy and Phenomenological Research*, vol. 94, no. 1, 2017, p. 121-143.

Rorty, Amelie, “The Deceptive-Self: Liars, Layers, and Liars2,” in McLaughlin, B., and A. Rorty (eds.), 1988, p. 11-28.

Schroeder, Mark, “Stakes, Withholding, and Pragmatic Encroachment on Knowledge,” *Philosophical Studies*, vol. 160, no. 2, 2012, p. 265–285.

Scott-Kakures, Dion, “At ‘Permanent Risk’: Reasoning and Self-Knowledge in Self-Deception,” *Philosophy and Phenomenological Research*, vol. 65, no. 3, 2002, p. 576-603.

———, “Can You Succeed in Intentionally Deceiving Yourself?” *Humana.Mente Journal of Philosophical Studies*, vol. 5, no. 20, 2012, p.17-39.

Stanley, Jason, *Knowledge and Practical Interests*, Oxford, Oxford University Press, 2005.

Stanley, Jason and John Hawthorne, “Knowledge and Action,” *Journal of Philosophy*, vol. 105, no. 10, 2008, p. 571-590.

Szabados, Béla, “Wishful Thinking and Self-Deception,” *Analysis*, vol. 33, no. 6, 1973, p. 201-205.

Talbott, William J., “Intentional Self-Deception in a Single Coherent Self,” *Philosophy and Phenomenological Research,* vol. 55, no. 1, 1995, p. 27-74.

Trope, Yaacov, and Akiva Liberman, “Social Hypothesis Testing: Cognitive and Motivational Mechanisms,” in E. Higgins, and A. Kruglanski (eds.), *Social Psychology Handbook of Basic Principles*, New York, 1996.

Van Leeuwen, Neil, “The Product of Self-Deception,” *Erkenntnis*, vol. 67, no. 3, 2007, p. 419-437.