7-2020

When Down Looks Like Up: Self-Deceptive Self-Handicapping

Kyle T. Hallam
University of Arkansas, Fayetteville

When Down Looks Like Up: Self-Deceptive Self-Handicapping

A thesis submitted in partial fulfillment
of the requirements for the degree of
Master of Arts in Philosophy

by

Kyle T. Hallam
University of Arkansas
Bachelor of Arts in Philosophy, 2015

July 2020
University of Arkansas

This thesis is approved for recommendation to the Graduate Council

Eric Funkhouser, Ph.D.
Thesis Director

Warren Herold, Ph.D. Committee Member	David Barrett, Ph.D. Committee Member

Abstract

In this thesis, I present a novel example of intentional self-deception as embodied in self-handicapping behavior. Self-handicapping is the proactive construction or acquisition of some obstacle to success in some domain, and is employed by individuals primarily as a means of deflecting blame for a failure or negative outcome. I argue that this behavior stands in a mutual, symbiotic relationship to self-deception. On the one hand, self-handicapping is the behavioral instantiation of the biased evidence manipulation which facilitates self-deception; while on the other hand, self-handicapping effectively functions to bias judgments in this way only in case concurrent self-deception sustains the behavioral process. If my account is accurate, the findings support Intentionalist theories of self-deception, broadly, but also highlights behavioral self-handicapping as a phenomenon worthy of philosophical attention.

Acknowledgements

I would like to thank a few of the contributors to this project, without whom it would not have been possible.

First and foremost, I want to express my deepest appreciation of Dr. Eric Funkhouser for offering his intellectual and professional guidance not only in the development of this work, but over of the course of my academic career. The thoughts which follow are unrecognizable from the inchoate formulations developed over a year ago, and without his efforts and insight this thesis simply would not exist.

I would like to thank the whole of the Philosophy Department at the University of Arkansas for fostering an atmosphere of curiosity and receptivity. Being a member of our program is a privilege for which I will be forever grateful.

Further, thanks the kind folks at the Southwestern Society of Mind for allowing me to present a brief version of this work at the 1st annual meeting. Without that forum, and the society's thoughtful feedback, the core arguments supporting this account would lack the vital foundation provided by the conference.

Lastly, to my family: words cannot express how valuable your love and support has always been. The final few months dedicated to completing this project consumed so much of your efforts in addition to mine. I did not complete this project alone, but only with you by my side. Thank you for never allowing me to sell myself short; I owe you my world.

Dedication

To Mom & Dad, for endlessly nourishing my spiritual and intellectual passions. I love you.

Table of Contents

Epigraph

"To be sure, I knew my failings and regretted them. Yet I continued to forget them with a rather meritorious obstinacy. The prosecution of others, on the contrary, went on constantly in my heart. Of course—does that shock you? Maybe you think it's not logical? But the question is not to remain logical. The question is to slip through and, above all—yes, above all, the question is to elude judgment. I'm not saying to avoid punishment, for punishment without judgment is bearable. It has a name, besides, that guarantees our innocence: it is called *misfortune*."

-Jean-Baptiste Clement-

Camus, *The Fall*

Section One – Introduction:

Self-Deception and the Neglect of Self-Handicapping

The contemporary philosophy of self-deception is primarily fueled by two disputes which arise from the "lexical" (Mele, 1997) approach to conceptualizing the phenomenon. On the lexical model, philosophers have looked to the meanings of the composite terms, "self" and "deception," investigating whether, and to what degree, self-deception is analogous to its interpersonal (other-oriented) counterpart. This formulation is by and large congruent with our natural language use and conventional understanding of "self-deception," but even so is nonetheless potentially subject to some conceptual opacity or incoherence.

Consider, if we pause to reflect on this conventional understanding of paradigm cases, we find two "apparent truths" (Van Leeuwen, 2013) about deception which are not easily applied to its reflexive counterpart:

(1) When deceiving another, I believe that *p* yet manipulate the target into falsely believing not-*p*, and the process of deception is terminated when the target comes to form that belief (not-*p*). Since in *self*-deception, though, I am both the perpetrator and the victim, it seemingly follows that I must believe that *p* while convincing myself, and coming to form the belief, that not-*p.*

(2) When deceiving another, as the perpetrator I intend to bring about the target's false belief; conventionally, deception is not typically taken to be an accidental occurrence. Again, since in *self*-deception, I am both the perpetrator and the victim, it seemingly follows that I must intend to bring about my own false belief.

Apparent truth (1) illustrates the "static" paradox, and (2) the "dynamic" paradox (Mele, 1997; 2001). We see that on the lexical model, it (paradoxically) follows from (1) that *S* (the self-deceived) holds contradictory beliefs (that *p* and that not-*p*), and from (2) that S acts so as to form a belief which they know to be false. Now, Mele (1997) abandons the lexical approach in light of these paradoxes, instead framing the issue as "essentially about explanation" (p. 93). That is, rather than conceptually constraining our understanding of self-deception in accordance with the semantic meaning of the terms, the explanatory approach looks to paradigm cases of irrational belief formation and "postulate[s] self-deception in particular cases to explain behavioral data…. And we should ask how self-deception is likely to be constituted…if it helps explain the relevant data" (p. 93). But, there is nonetheless still mystery surrounding the doxastic and motivational state(s) which accompanies self-deception, and on these grounds, the static and dynamic puzzles are not dissolved merely by altering our approach. Basically, the fundamental curiosities remain: what does S believe in self-deception, and how does S acquire that belief?

Part of my aim in this work is to bring some clarity to this conceptual confusion, by way of illustrating an empirically demonstrable type of action or pattern of behavior that is quite similar to, and influential upon, self-deception, but one which has not received adequate attention in the philosophical literature. Specifically, I argue that the common self-protective behavior known as "self-handicapping" stands in a symbiotic, mutually reinforcing relation to self-deception. I will focus on *behavioral* self-handicapping, S's active construction or acquisition of an impediment to their success in a given task, arguing that we engage in this behavior in order to (deceptively) acquire or retain beliefs featured in our desired or ideal self-concept. In some cases, S may merely claim or avow such an impediment ("I was sick, so I had to miss the meeting"); but in the cases at hand, S actually constructs such an obstacle (say, by drinking to induce a hangover, thus making themselves sick).

As I have in mind, the process unfolds roughly as such: S is placed into a situation in which they expected to be evaluated along some dimension which is vital to their self-identity or self-esteem; S feels as though they will be *judged*. In the face of this evaluation, the prospect of failure brings with it a threat to S's self-perceived identity, which, in turn, motivates S to protect that valued identity. So, in order to prevent the possibility of a neutral and objective evaluation, S constructs or acquires an obstacle that may be used to explain away the negative outcome, deflecting blame from S and onto that source. However, as we will see, this works both ways, since in the case of success, S can claim victory *in spite of* the acquired obstacle, which suggests an innate talent or virtue strong enough to overcome that disadvantage. Thus, in these cases, self-handicapping serves to strengthen or enhance S's desired self-identity.

For a quick example, imagine the student whose self-identity is importantly dependent on a self-belief in their innate intellectual superiority to others. However, suppose that as final exams approach, S becomes a bit anxious, since a poor grade would challenge these beliefs and potentially threaten S's identity. So, during finals week, S elects to party with friends or to play video games on the nights before big exams, even though these behaviors are not most conducive to success, indeed making failure more likely. But, S knows unconsciously that they may blame any bad grade on this lack of preparation. Now, this behavior may indicate S's lack of prudence, for which they may deserve blame and which may reflect poorly on their character, but crucially, it does not warrant any conclusions about S's trait intelligence. Thus, the behavior functions to shield valued features of S's identity from the threat of a negative evaluation. Suppose though, that S succeeds in the task, in this case getting an 'A' on the exam. Then, S may claim (and believe) something like, "See, I didn't even have to *try* to pass. I'm just that smart," solidifying S's self-perception as innately intelligent.   Actors in these cases create a kind of *win-win* for themselves, since the external impediment might function to either externalize blame or to internalize success, depending on the outcome.

From this rough description, perhaps the reader is familiar with the behavior, and acquainted with a few folks apt to engage in it; self-handicapping is quite ubiquitous. Now, I take it that our conventional understanding of self-handicapping is one in which the behavior is designed to save-face in the presence of an audience whom S desires to impress. The student above may self-handicap so that others believe them to be superior, whether S comes to believe this themselves or not. However, as we will see, these behaviors are not always (or even most often) directed at an external audience. Instead, behavioral self-handicapping can function as a method by which we manipulate ourselves into forming or retaining desired beliefs; a method by which we self-deceive. Specifically, I argue that in self-deceptive self-handicapping, S intentionally engages in self-handicapping behavior as a method of self-*deception*, in the active sense of deceiving oneself. Although, I argue that self-handicapping cannot effectively function as such unless S is unaware about their intentions in self-handicapping and the purpose for that behavior.

In what follows, then, I provide a conceptual overview of the philosophy of self-deception, generally, then more narrowly focusing on the dynamic problem. In the contemporary literature, proposed answers to the dynamic paradox fall into two camps: *deflationary* (Motivationalist) or *Intentionalist* (traditionalist) accounts. I will argue in support of Intentionalism, relying on empirical demonstrations of self-handicapping behavior to show that agents in the relevant cases are sufficiently motivated to self-deceive, and that these motivations culminate in intentional self-handicapping behavior, which is guided by the goal of self-deception and effectively serves that end. If my account is accurate, we find some support for Intentionalism generally, but this work also hopes to highlight self-handicapping behavior as worthy of attention in contemporary discussions of self-deception. As we will find, while largely ignored by philosophers, this behavior is intimately connected to self-deception, as well as other philosophically important concepts like intentionality and practical reasoning.

Section Two: The Static and Dynamic Paradoxes of Self-Deception

Mele (1997; 2001) is responsible for a great deal of the progress in the philosophy of self-deception. Mele's deflationary view is influenced by his rejection of both of the assumptions housed in the lexical model, arguing instead that standard deception requires neither any concurrent beliefs between the deceiver and deceived, nor does it entail any intention to deceive on part of the perpetrator. In rejecting the lexical model, Mele shifts the primary method of investigation to one that is focused on the relationship between affective states, motivation, and cognition, without necessarily attempting to salvage the analogy to standard deception.

One effective way to understand this framework is by an appeal to cognitive biases, and especially to the distinction between "cold" vs. "hot" (*motivated*) biases (Kunda, 1990). When S engages in *cold*-biasing, the acquisition of a false belief is in some sense innocent, as the biasing is accidental – S has no immediate goal or desire which explains the biased cognition. "Hot," or motivated biases, on the other hand, are judgments which are slanted in light of some goal, or because we have a stake in the truth-value of the relevant proposition. With respect to the requisite motivation, S's goal may sometimes be just that of forming an accurate judgment, in which case we should not necessarily expect biasing. However, sometimes we engage in hot biasing so as to reach a "particular, directional conclusion," (Kunda, p. 480), often as a method of self-protection and enhancement, or because the biased judgment facilitates some ulterior or supervening goal. Stock examples of hot cognition include bias in recalling beliefs or memories (p. 483) and the biased selection and evaluation of sources of evidence (p 489). Self-deception, by these lights, is at least an instance of motivationally biased evidence manipulation, since self-deception "is not simply a cognitive mistake; it is a *motivated* misrepresentation" (emphasis original) (Funkhouser, 2019, p. 54).

For instance, recall the previous example of the student who possesses an inflated sense of native intellectual ability. If this individual fails a big test while the rest of the class does quite well, the student may be motivated to attribute their own failure to a one-off poor performance, avoiding unwelcomed evidence to the contrary, while perceiving the success of others to be simply a *fluke*. Here, S is motivated to avoid admitting the possibility that their classmates perhaps possess superior intelligence, and is therefore motivationally biased in seeking out, evaluating, and rationalizing evidence, so as to avoid the unwelcomed belief and retain the welcomed one. Thus, the acquisition of false beliefs consequent to hot biasing is not likewise innocent and accidental as are cases of cold biasing, since it is not merely a mistake in reasoning, but something self-serving which triggers the biased judgment.

The hot/cold cognition distinction significantly informs the contemporary philosophy of self-deception as a widely agreed upon "minimal conception" of self-deception (Funkhouser, p. 54). Still, there are various ways in which motivational bias may manifest as deviant cognition or behavior, and self-deception is supposed to be conceptually distinct from other potential motivationally biased judgment. For instance, the self-important student may display thorough irrationality in excusing her poor test scores, but we can nonetheless doubt whether from this alone we should conclude that the student is properly in self-deception.[1] It is logically possible that all instances of self-deception may be accounted for by motivational bias, but not all instances of motivational bias constitute self-deception. The latter is a more robust cognitive deviance, and often suggests the presence of a more narrowly focused motivation. Whatever the

[1] For instance, (Scott-Kakures, 2002) argues that wishful thinking can be distinguished from self-deception in that the latter, but not the former, involves reflective reasoning and an error in self-knowledge. Conversely, (Lynch, 2016) argues that willful ignorance is necessarily intentional, while self-deception is "typically" not (p. 521).

potential conceptual nuances, recent philosophy in self-deception has devoted itself to making sense of the process and product of this biased judgment, and under this framework, we find substantially less effort involved in resolving the lexical paradoxes, and a great deal more emphasis on explaining stock examples of self-deception. While I do not adopt the lexical model in this work, the language of the static/dynamic "puzzle" is nonetheless helpful, since, as previously stated, the fundamental questions remain: what is the cognitive process involved in, and what is the product of, self-deception?

The Static Puzzle: The Deflationary Position:

Mele (2001) offers the most prominent Motivationalist account, which aims to *deflate* the puzzles, showing the paradoxes to be illusory, and explaining self-deception as a relatively uncomplicated instance of motivated irrationality. Mele rejects the lexical constraints on dual belief and intent, and in all, the following four conditions are jointly sufficient for an instance of self-deception (pp. 50-51):

> "1. The belief that *p* which *S* acquires is false.
> 2. *S* treats data relevant, or at least seemingly relevant, to the truth value of *p* in a motivationally biased way.
> 3. This biased treatment is a nondeviant cause of *S*'s acquiring the belief that *p*.
> 4. The body of data possessed by *S* at the time provides greater warrant for ~*p* than for *p*."

As it concerns the doxastic paradox, (1) is the only directly relevant condition. The other conditions bear on S's methods of evidence manipulation (2), how those methods shape S's judgment (3), and the external evidence available to S (4). Mele argues that none of these conditions are individually necessary, and argues positively that (4) is plainly *not* necessary. But it nonetheless follows from this account that condition (1) is necessary for instances in which the person is self-deceived *in believing* that *p*, since "a person is, by definition, *deceived in* believing that *p* only if *p* is *false*" (emphasis original). However, this is a "purely lexical point…and in no way implies that the falsity of *p* has special importance in the *dynamics* of self-deception," on Mele's account (2001, p. 51).

Since, according to Mele, the doxastic end-state of self-deception is simply false belief, we should pause to take note that his account does not argue that in self-deception S *replaces*, *represses*, or *sublimates* an existing belief (Funkhouser, 2019, p. 88). It is here that Mele departs from the lexical model in his rejection of the dual-belief condition, that S both believes *p* and ~*p*. The deflationary position hopes to sidestep the static puzzle by arguing it is sufficient for an instance of self-deception that S acquire a false belief[2], in turn eliminating any conceptually necessary doxastic conflict or contradiction between S's belief that *p* and other antecedent or concurrent beliefs. Therefore, it follows from Mele's account that self-deceptive belief is not inherently paradoxical, since S's belief that *p* need not conflict with other doxastic states.

However, one may question both the necessity and sufficiency of false belief as the (sole) doxastic attitude properly belonging to self-deception. For instance, a common line of objections argue that Mele's account neglects the anxiety or *tension* which accompanies the self-deceptive venture and/or the final doxastic product (Audi, 1982; Bach, 1997; Barnes, 1997; Funkhouser, 2005). Audi, for example, argues that in self-deception, S unconsciously knows the truth (not-*p*) but "sincerely avows, or is disposed to sincerely avow," belief in its negation, that *p* (1982, p. 135). Self-deception, on this reading, does not terminate in the acquisition of a false belief, since S unconsciously retains a hold on the truth, while merely claiming the deceptive belief. Self-deception, then, is essentially characterized by this tension or dissonance (p. 137). Somewhat

[2] Of course, S must acquire this belief in the proper, self-deceptive way, but this is a matter about the dynamics of deception, however, and not wholly relevant to the static puzzle.

similarly, Barnes argues that self-deception is motivated by an “anxious desire,”[3] and that the process “always involves some reduction of anxiety,” (1997, p. 35).[4] And, lastly, Bach and Funkhouser, respectively, argue that S holds the truth “dangerously close to hand” in self-deception (Bach, p. 105), which renders the process one of ongoing epistemic manipulation. Then, throughout this deceptive process, S’s strategic avoidance of and attention to evidence is good reason to believe some such tension is present (Funkhouser, pp. 300-301).

Mele has no qualms in admitting that some instances of self-deception may (contingently) involve anxiety (reduction), but describes this state as an affective or emotive anxiety, not the cognitive anxiety we associate with racing thoughts (Mele, 2001, p. 55), and in any event is not a necessary condition on his model. Whether self-deceptions entails anxiety is an empirical question, but for present purposes in reviewing the literature, however, we need only note that the Motivationalist position is not committed to the existence of dual, contradictory beliefs in self-deception. In all, the deflationary position hopes to avoid paradox by simplifying our understanding of the nature of the mental state(s) involved in self-deception. Certainly, the simplicity of the deflationary position is a virtue, since it allows for an explanation of self-deception which relies only on a singular false belief, dissolving the potential paradox; but what these accounts afford in coherence, they may sacrifice in explanatory force.

For, if the false belief that *p* is sufficient to account for the doxastic state produced by self-deception, then accounts following suit may struggle to distinguish the belief-state involved in self-deception as compared to that involved in other instances of motivated bias. However, as

[3] There is no need to detail which features of a desire are sufficient to render it an “anxious” one, since it seems clear that most desires do not have this property. My basic hedonistic desire to eat a cookie, for instance, is not motivated by anxiety.

[4] For a more thorough review of Barnes’s position, see *Seeing Through Self-Deception* (1997, Ch. 2& 3).

we saw, while some have argued that self-deception is distinct on grounds of an accompanying state pf anxiety, I am unwilling to claim that this condition is necessary. To posit the necessity of anxiety in self-deception, one must then deny the phenomenon in all instances of motivated irrationality in which S is "free from psychic conflict," (Mele, 2001, p. 53). This is too stringent a condition; surely there are *some* instances of conflict-free self-deception. Furthermore, even in case such anxiety is necessarily present in instances of self-deception, this attitude is neither the sole feature which distinguishes self-deception, nor that feature which best illustrates and explains it.

So, the outcome of this dispute is not sufficient to determine our broader understanding of self-deception. Simultaneously, though, it seems *prima facie* apparent to me that a significant number of (if not most) instances of paradigm self-deception, *do* involve anxiety reduction mentioned by Barnes, or the sensitivity to the truth as suggested by Audi and Bach. So, I will not take a firm stance with respect to the precise role of doubt, anxiety, or tension in deceptive belief. I do not think such a condition is a necessary feature of self-deception, though accounts which outright neglect the common experience of doxastic tension will tend to fall short in providing a robust explanation of the phenomenon. In accordance with Mele's view, then, from this point forward this paper accepts the sufficiency of false belief in accounting for the self-deceptive state, but I suspect such instances constitute the exception, rather than the rule.

Psychological Partitioning

Partitioning, or *divided mind*, accounts meet the dual belief condition head on, though not attempting to *deflate* the paradox, but instead deny the paradox ostensibly inherent to the dual belief condition. Psychological partitioning theories (Pears, 1984; Rorty, 1994; Davidson, 2004) appeal to unconscious mental states and cognitions in order to account for dual beliefs, positing a mind in which contradictory beliefs, values, and desires reside with quasi-independent subpersonal operators. So, the fundamental claim shared by these accounts of self-deception is that the dual-belief condition is not a paradox, since one need not argue that the entity doing the deceiving is identical to that which ends up deceived.

Unfortunately, I doubt that these descriptions of the self will prove any less abstract as we progress, but I think we can make sense of these claims, even if we find ourselves unconvinced. Beginning with Rorty (1994), this account offers a quite Freudian description of self, arguing that this *self* (or ego) is composed of various subconscious, subpersonal entities, and moreover, that these entities may possess unique motivations and beliefs. On this reading, individual persons are some amalgamation of, and the interaction between, "sub-systems……whose interaction is only precariously integrated" (pp. 223-224). Most of all, however, the entities are capable of engaging in "nonpurposive but intentional operations" (p. 224) toward the deceptive end. Accordingly, this account hopes to avoid paradox by arguing that the system or agent *doing* the deceiving is not identical to that which is *deceived*, even though both factor into a singular identity. What is most striking about this account is the agential nature of the subpersonal units, and their ability to deceive, and their liability to fall into deception; but whether such agency is present is not particularly prescient to the problem at hand. As a matter purely of S's doxastic attitude, on this reading, the upshot is that some subconscious, subpersonal unit actively suppresses or conceals some information from the feature with which S identifies in the context. Unlike Mele, Rorty argues that self-deception need not produce a false belief, instead arguing that the process may result in S's acquiring a *true* belief, or maybe no belief at all (p. 216).

Similarly, Pears (1984) describes self-deception as a kind of psychological "schism" (p. 74). Again appealing to sub-personal entities, this account argues that composite features of the self contend with one another for identification with the “main system” – S’s coherent, composite self (p. 101). When sufficiently motivated, Pears claims, the subsystems drive the main system to selectively attend to preferred evidence and suppress unwelcomed evidence, thus acting as agents in manipulating S’s (the main system) beliefs. Rorty likened these subsystems to the “President’s Cabinet” (p. 217), since each agent works both independently and in concert with the others to maintain and manipulate the functioning of the primary agent, and I think this analogy is apt for Pear’s account, as well. Both explanations of self-deception describe self-deception as something of a resolution to an internal conflict between competing agents, vying for identification.

Admittedly, this language is psychologically rich, but upon reflection, I suspect that we are familiar with the description of an individual experiencing conflict between their “identity as a parent” and “identity as an employee,” for example. In this sense, the self is more readily understood as totality of our values, beliefs, affiliations, and experiences. It also seems clear that we do differentially identify with these various features, and that these motivate us in often conflicting directions. Though, while this much seems uncontroversial, to assume that these constituent features of the self may also possess individual beliefs as is the case with motivations, is simply to beg the question. So, ultimately whether the concept of self as is properly described as the composition and interplay between competing sub-agents, and moreover, whether such a notion could withstand empirical scrutiny, are issues which cannot be sufficiently addressed here. What Rorty and Pears highlight in their work, though, is the experience of our being conflicted in motivation and belief, and the (at least) superficial appearance of some agency in the self-deceptive project. Now, the question of agency demands its own discussion, but these accounts address the static puzzle by appeal to various entities which may possess the conflicted beliefs, in turn arguing that these entities are divided mind accounts are vital in advancing the Intentionalist position.

So, while Rorty and Pears are on to something in arguing for the explanatory importance of the seemingly purposive[5] behavior in self-deception, it is nonetheless fair to question the degree to which this apparent agency is explained by the subpersonal agents described. These theories ask more of us than for a mere admission of unconscious belief or motivation into the account of self-deception, indeed going further in requiring that our ontology includes some quasi-autonomous *entities* which compose a coherent self. One immediately wonders whether such a postulation is empirically verifiable, if not plainly superfluous on conceptual grounds.[6]

With this in mind then, looking to Davidson (2004) as an example of a (unreservedly Intentionalist) partitioning theory in which the mental sequestration of conflicting states does more work in explaining the sustained *state* of self-deception, as compared with Pears' work which is most basically interested in the deceptive *process*. Specifically, Davidson is fundamentally concerned with irrational doxastic states associated with self-deception, and the accounts breaks from Rorty's and Pears' in claiming that self-deception should not be imagined as "two minds each somehow able to act like an independent agent; [but] rather that of a single mind not wholly integrated; a brain suffering a perhaps temporary lobotomy" (p. 221). Davidson argues that the partition he posits is simply the "metaphorical wall" functioning to suppress at least one of the conflicting beliefs in order to prevent both beliefs from being "destroyed" (p. 220). This metaphor is not intended to constitute a full-fledged psychological explanation of self-deception, but to stand in for whatever the psychological mechanisms are which facilitate

[5] I reserve any language of "intention" in this section. As mentioned, the issue of intention will receive a focused discussion in what follows.

[6] As a deflationary position might argue, for instance in (Mele, 2001, Ch. 2&3).

belief suppression as such. That is all a basic partitioning account requires – that there exists some psychological mechanism by which our motivations and beliefs are sometimes insulated from other mental states, or by which these beliefs are otherwise rendered inaccessible. As stated previously, my work here is not meant to offer any robust analysis of the doxastic states consequent to self-deception, but a grasp on the divided mind accounts is nonetheless prerequisite to an understanding of the Intentionalist theories upon which my analysis of self-deceptive self-handicapping ultimately relies. So, the works just outlined inform my arguments to follow in a few ways.

First, I accept Mele's claim that S's false belief that *p* is sufficient to account for the doxastic state associated with self-deception. Though, I am nonetheless confident that many instances of self-deception induce a doxastic state not entirely consumed by the folk concept of belief, or perhaps even a state not accurately described as a belief at all. So, a thorough account will likely find that paradigmatic instances of self-deception involve some mental state in addition to, or as distinct from, a false belief. Moreover, as I hope to show, Rorty's and Pears' works prove valuable in accounting for the empirical data related to self-handicapping behavior, since in these cases, the motivation to engage in self-deceptive self-handicapping is sourced from relatively specific features of S's self-perceived identity. Self-deceptive self-handicapping is narrowly focused in a manner perhaps unlike the deceptive processes involved in some *garden variety* cases; as motivated by, and serving to protect, specific features of the self, the language of subpersonal agents with (quasi-)independent goals, desires, and beliefs is all the more prescient. However, if it turns out that the language employed by Rorty and Pears is inaccurate, such a finding does not significantly undermine my account, as ultimately, we can make do with

Davidson's metaphor representing beliefs and desires as psychologically *walled off* from one another.

Before proceeding with this foundation, though, diligence requires that I mention here an alternative class of "revisionary" accounts (Funkhouser, 2009; Baghramian, 2013) which argue that folk psychological concepts like desire and belief cannot provide an adequate answer to the doxastic puzzle.[7] Now, certainly the validity of our lay understandings of mental phenomena are often strained by the apparently paradoxical, and at least demonstrably irrational, cognitions and behaviors appurtenant to self-deception. But, my account need not adopt any specific theory of mind in order to proceed. Instead, we may adopt a rather broad understanding of belief as simply "the psychological attitude in which a person regards a proposition as true…. [T]o believe a proposition is to regard it as true all things (i.e., psychologically relevant factors) considered." (Funkhouser, 2009, p. 5). In adopting this broader view of belief, my account is therefore potentially acquiescent to various theories of mind or descriptions of doxastic attitudes. Having outlined the responses to the static puzzle, and having positioned this account somewhere within the family of partitioning accounts, we can now turn to the dynamic puzzle – that related to S's agency and intention in carrying out the self-deceptive task.

The Dynamic Paradox: Is SD Intentional?

Mele (2001) succinctly expresses the dynamic paradox, and the responsibilities of philosophers who attempt to resolve it, as follows:

[7] "[T]here are other cases in which the self-deceived produce a confused belief-like condition, so that it is genuinely indeterminate what they believe with respect to p. Yet, such people, the deeply conflicted, can be wholly self-deceived nonetheless. If this is true, then self-deception does not produce a common product…. The protracted debates over the motives and products of self-deception are largely due to disputants using folk-psychological tools that simply are not up to the task" (Funkhouser, 2009, p. 13).

> If I am trying to bring it about that I believe that I am a good driver…perhaps by ignoring or downplaying evidence that I am an inferior driver while searching for evidence of my having superior driving skills—won't I see that the "grounds" for belief that I arrive at in this way are illegitimate? And won't I therefore find myself still lacking the belief that I am a good driver? A predictable reply is that the…efforts…are not conscious efforts and therefore need not stand in the way of their own success in the way just envisioned. Whether, and to what extent, we should postulate unconscious tryings…depends on what the alternatives are. (pp. 13-14)

Paradigmatic instances of self-deception are motivated by the desire for *p* to be true of the world (Mele, 1997; 2001), or alternatively, to simply *believe* that *p* is true regardless of the material reality (Nelkin, 2002; Funkhouser, 2005).[8] Typically, in explaining the goal-oriented actions of others, we rely on a standard belief-desire model, positing the existence of an intention to link the individual's mental states with the consequent behavior designed to satiate the desire or achieve the goal. However, in the case of self-deception, the supposition of an intention to link the desires with actions may render such an explanation incoherent. For, if my intention is to deceive myself, then won't I recognize that my belief that *p* is formed on epistemically shaky grounds, thus preventing me from actually acquiring that belief? It seems, at least initially, that one's intention to self-deceive would render the entire goal-pursuit futile.

Answers to this challenge are, broadly, classified as either Intentionalist (agency) views, or deflationary (anti-agency) views. Intentionalist accounts claim that in paradigmatic instances of self-deception, S's desire that *p* (or to *believe* that *p*) produces an intention to deceive oneself into believing that *p*. On the other hand, standard deflationary accounts do not *a priori* reject the possibility of intentional self-deception, but rather argue that such examples are fringe instances,if at all demonstrated in the literature. So, in all, the disagreement is rooted in

[8] "World-directed motive" vs. "Mind-directed motive" (Funkhouser, 2019, p. 66)

explanatory superiority, where each camp claims to explain the most common instances of self-deception.

The Deflationary Position: Eliminating Intention

The deflationary position generally denies that intention plays a significant role in standard cases of self-deception, arguing that affective states or other motivations are causally responsible for S's coming to believe that *p* without an accompanying intent to deceive. As we have seen, Mele (1997; 2001) argues that the following four conditions are jointly sufficient to constitute an instance of self-deception (pp. 50-51):

> 1. The belief that *p* which *S* acquires is false.
> 2. *S* treats data relevant…to the truth value of *p* in a motivationally biased way.
> 3. This biased treatment is a nondeviant cause of *S*'s acquiring the belief that *p*.
> 4. The body of data possessed by *S* at the time provides greater warrant for ~*p* than for *p*.

(2) and (3) are uniquely related to the dynamic puzzle, since these call for an explanation of how motivation can bias one's cognition, what that biased treatment of evidence looks like, and what it means for that treatment to be a "nondeviant" cause of S's belief that *p* with an appeal to the intention to self-deceive.

According to Mele's model, S desires that *p* is true of the world, and this desire influences the selection and evaluation of hypotheses pertinent to forming a judgment about whether *p*. Specifically, Mele argues that S's hypothesis testing, and the resultant acquisition of a belief, is mediated by the costs associated with *falsely* believing that *p*. Mele argues that the acquisition of a false belief entails some costs—in terms of affect, social value, or "whatever one might expect" from forming a belief (Mele, 2001, p. 35). When the false belief that *p* is costly, S ought to be influenced to test disconfirming hypotheses, and interpret the data in a light conducive to the rejection of *p*. Alternatively, high costs associated with the false rejection of *p* should motivate the search for confirmatory evidence. In the case of self-deception, when S's acceptance of the false belief that *p* is less costly than the false rejection of the truth, that not-*p*, Mele's model predicts self-deception. That is, in all cases of self-deception, S's false acceptance of the proposition entails the concomitant rejection of the true proposition that not-*p*, and when the false acceptance is less costly (or roughly equally costly) than the accompanying false rejection of the truth, S should be inclined to self-deceive (given the requisite desire).

Now, all this is quite abstract, so we can look to the following example to help make some sense (Mele, 2001, pp. 37-38).

> Bob wants it to be true that he is the best third baseman in his league. Owing partly to that desire, he has a lower threshold for believing that he is than for believing that he is not. Bob examines the statistics on the competition and decides, correctly, that his main competitor is Carl. Bob and Carl have the same fielding percentage, but Carl has a few more home runs…several more runs batted in…and a higher batting average. However, Carl's team is much better than Bob's, and, as Bob knows, players on better teams tend to have more opportunities to bat in runs…and to hit home runs…Bob takes all this and more into account and comes to believe that he is a better player than Carl.
>
> As it turns out, however…Carl is the better player…. Carl's team was far superior to Bob's, but…Carl batted many fewer times than Bob…And …given this statistic, Carl still outperformed Bob in home runs and runs batted in.
>
> …Bob's coming to the conclusion he does is explained in significant part by his having a lower threshold for believing that he is better than Carl than for believing that this is not so…. Given the difference in thresholds, Bob's acquiring the belief that he is the better player requires less evidential support than does his acquiring the belief that this is not the case.… And there is no evident need to suppose that Bob tried to bring it about that he believed, or tried to make it easier for himself to believe, that he is the better player, or tried to reduce the probability that he would believe that Carl is the better player.

In all, the costs for Bob of falsely rejecting the true belief that Bob is the superior player do not outweigh the costs associated with his falsely believing himself to be the superior player. For Bob, the costs of rejecting the truth about his relative talents (dissatisfaction of his desire) may be higher than the costs of false acceptance of the false belief (whatever costs are associated with

Bob's being deceived), and in such a case, Mele predicts self-deception. Fundamentally, according to Mele, self-deception is explained on the basis of our tendency to avoid costly errors in belief.

We must take note that the deflationary position does not deny the "unproblematic possibility" of "intentionally deceiving oneself," as embodied in an intentional action which is causally responsible for S's belief that *p* (Mele 1997, p. 99). But notice here that the intention in action does not adequately explain why it is that S came to believe that *p*. For example, I may intentionally walk across the room, consequently stepping on my dog's toy and experiencing some pain as a result. While my walking was intentional, I did not likewise intend to step on the toy nor bring about my experience of pain, so the outcome of my actions is not *explained* as a function of my intention but as an accidental byproduct.[9] So, the deflationary position argues that S's intentions do not explain, in this way, the acquisition of the self-deceptive belief – that we can "only—or best—explain" self-deception by appeal to intention (1997, p. 99). So, the primary burden of Intentionalist accounts is illustrating explanatory superiority.

One feature of self-deception that motivates the Intentionalist position is the potential insufficiency of desire to account for the "selective" nature of self-deception (Bermudez, 1997, 2001). That is, one may desire various things in various contexts, but typically, these desires alone are insufficient to render an acquisition of the false belief. For instance, I may desire it to be the case that I have possess the talent to play in the NBA (or the desire to *believe* as much), but I do not in turn deceive myself about the truth of the matter. Moreover, this seems to be the typical case; I have many desires, the majority of which, however, do not produce self-deceptive belief. It is not entirely clear how desire is functioning according to the deflationary account,

[9] Audi (1982) argues that self-deception is explained as this kind of action and outcome.

since we most often posit the existence of some intention to bridge gap between motivation and action.

Now, Mele appeals to “internal biasing” (2001, p. 60), as the biased interpretation of available evidence, as one mechanism responsible for self-deceptive belief acquisition, and such rationalization surely does corrupt our beliefs. Further, I do not claim that these kind of internal rationalizations are intentional in the relevant sense. However, in self-deception, S is not merely a passive recipient of evidence, which is in turn rationalized, but rather seeks or avoids certain evidence, so as to acquire the desired belief; Bob attended to certain statistics and avoided unfavorable ones. While the deflationary position may account for S’s deceptive internal biasing once they are in the presence of evidence, in self-deception S is not merely the passive recipient of evidence, instead selectively engaging with sources which contribute to the deceptive venture. Mele (p. 60-61) notes that self-deception may rely on strategic “input-control”—the somewhat tactical selection and avoidance of (dis-)favorable evidence—but, the Intentionalist charge is that desire or motivation, alone, is insufficient to explain this behavior. While I will grant Mele’s claims regarding internal biasing, it seems that desire is performing a unique function in self-deceptive input control, which we rarely if ever witness in other behaviors.

Intentionalism resolves this potential pitfall, purporting to explain S’s behavior by way of an intention to bring about the belief that *p*, linking the motivation to believe with active evidence manipulation. However, these accounts face a similar worry, since just as desire is perhaps insufficient to explain self-deception, desire alone is insufficient to explain the formation of an *intention* to self-deceive. Thus, an adequate theory of Intentionalism will also answer its end of the “selectivity problem,” explaining the manifestation, presence, and function of the requisite intention (Jurjako, 2013; Funkhouser, 2019, pp. 112-114).

Generally speaking, very little of the Motivationalist position depends on a sophisticated understanding of ‘intention’. Mele, however, has quite a bit to say about intention elsewhere (1994; Adams & Mele, 1989), and, in his work on self-deception, refers to intention as an “executive attitude” that involves “making [one’s] mind up” about acting on some motivation (1997, p. 98). So, as one last detour before submitting some positive claims, a more thorough dissection of Mele’s thoughts on intention are helpful for making clear his position on self-deception, but also for helping to highlight the mental states and behaviors we ought to look for in ascribing intention in self-deception.

According to Mele & Moser (1994), intentions include both motivational and cognitive aspects. The motivational component, qua desire, contributes to the goal-formation and the initiation of action (p. 26). On the cognitive side, intentions function to provide a “plan” for S to bring about the satiation of their desire and to structure S’s actions accordingly. This much accords with the standard belief-desire model of intention, but Mele concludes that S has an intention only if S is “settled” on carrying out a plan of action in order to achieve a goal or bring about a desired state. The action initiation and goal-pursuit sparked by an intention results in a complex process of behavior, such that intention functions in multiple ways, to “initiate” and “motivationally sustain” action; to “guide and monitor” action; to “coordinate activities”; and to “prompt and terminate practical reasoning” (p. 27). The Motivationalist position argues that this mental state, intent, is absent in the standard case of self-deception.

The crux of determining whether an action is intentional, then, relies on the degree to which S’s mental state is like that described in (Mele & Moser 1994), and whether this state sufficiently “guides and sustains” their action. This, in turn, depends on whether S is responsive to goal-directed feedback, adjusting their behavior in response to its effectiveness in attain the goal. Therefore, in advancing my claims, I will argue that in cases of self-deceptive self-handicapping, S is settled on self-deceiving[10], and moreover, that this disposition structures and guides S's self-handicapping behavior. In these cases, I argue that S responds to resources which assist in self-handicapping, adjusting their behaviors in accordance with these and other situational constraints, ultimately in a manner which facilitates acquisition of the self-deceptive belief. I will go on to argue that S intentionally engages in self-handicapping just because that behavior is conducive to self-deception, and further, that S is unconsciously aware of their motivations and the function of self-handicapping, strategically employing the behavior in light of the motivation to believe that *p*. First, however, we must pin down what we are talking about in referring to "self-handicapping".

Section Three: Self-Handicapping Behavior – Conceptualizing the Phenomenon

Self-handicapping is a specific type of *excuse-making*, and when the behavior becomes self-deceptive, we may describe the people in these cases as making an excuse for themselves. Generally, we offer excuses in hopes of deflecting some blame away from ourselves, focusing that attention on factors beyond our control, and I take it that most are fairly well acquainted with this understanding. Stated more formally, excuses function to "shift the causal attributions" for negative outcomes from "sources relatively central to one's sense of self to sources relatively less central…" (Snyder & Higgins, 1988). When we offer the standard excuse, we express that a given outcome is *not our fault*.

The motivation to offer excuses is triggered by evaluative situations, and specifically those in which S perceives the situation as one in which there is a risk of having their identity, personality, or intrinsic traits *linked* to – perceived as causally responsible for – some negative

[10] My account is therefore *mind-directed*.

outcome. When an individual expects to be evaluated by others in light of some course of action, the potential for the outcomes of these actions to be attributed to the individual's trait characteristics or intrinsic qualities becomes all the more salient. The stronger an individual's expectation, the more subjectively salient this possibility. If S expects the outcome of their actions to be positively evaluated, they may welcome such an attribution; but when the agent expects a negative outcome, we find that S is motivated to dissolve perceptions that something internal to them is responsible for the outcome.

Synder and Higgins (1988) distinguish between verbal (claimed) and non-verbal (behavioral) excuses (p. 243-244). Our standard understanding of what it means for one to make an excuse most often implicates the verbal type. These verbal excuses are ubiquitous: "I am sick"; "I overslept"; "I just forgot"; "I got distracted". In these cases, we see that the claims indeed function to shift attributions of responsibility; even though S may be blamed for oversleeping or forgetting, these (claimed) mistakes do not typically reveal anything native to the excuse-maker, but are instead instances of an ostensibly contingent, temporary deviance.

Behavioral excuses express basically the same message, and have basically the same function, as their verbal counterpart. But these instances uniquely involve some "physically observable manifestations" beyond, or in place of, the mere avowal (p. 243). For example, the authors ask us to imagine the starting pitcher of a baseball game who is having a poor performance. As the manager comes to remove the pitcher from the game (much earlier than anticipated), the pitcher begins to rub their throwing arm and wince in pain, leaving the audience and teammates to witness the pitcher's ostensible disadvantage. The effect, if successful, is not necessarily different than it would have been had the pitcher simply claimed the existence of arm pain, so the basic function is retained.

Snyder and Higgins further distinguish between types of excuses on the grounds of whether the excuse is *retrospective* (past-oriented) or *anticipatory* (future-oriented) (pp. 244-245). Retrospective excuses are likely most familiar, and these excuses function to explain away an outcome which has already occurred. These "after-the-fact" excuses hope to weaken causal attributions which have already been made, and are therefore, at best, a *make-up* for an already poor performance. Anticipatory excuses are forward-looking, which is to say they function to decrease the likelihood of an undesirable attribution *prior* to the relevant action and outcome. Therefore, anticipatory excuses provide a potential benefit not available to the retrospective excuse-maker, since these have the "capacity to structure the context within which evaluations take place so as to obscure the diagnosticity of evaluative feedback or preempt its utility altogether" (Berglas, 1989, p. 268). That is to say, anticipatory excuses, but not their retrospective counterpart, might function to proactively corrupt the perceived link between the actor and outcome prior to any attributions ever being made. In this way, anticipatory excuses create a sort of *win-win* situation, since in the case of failure S may explain the outcome away by appeal to the excuse, while in the case of success S accepts credit for the outcome *in spite of* that obstacle.

Putting it all together, self-handicapping can manifest either verbally or behaviorally, but the empirical literature is widely in agreement that self-handicapping is anticipatory in nature (Jones & Berglas, 1978; Higgins and Berglas, 1990; Hirt & McCrea, 2001; McCrea, 2008, 2012). So, self-handicapping is not an attempt to make up for a failure, but to proactively prevent the potential for blame prior to any performance. Formally, "self-handicapping refers to the strategic creation or claiming of barriers to one's own success prior to a performance in order to protect evaluations of the self" (McCrea, 2012, p. 76). These "barriers" to success serve the role of deflecting blame away from S, as is the case in standard excuse-making.

In verbal or claimed self-handicapping, S simply avows (whether sincerely or not) the existence of an external impediment to their success prior to performance. For instance, in investigating the claimed self-handicapping strategies of competitive athletes, studies have replicated the finding that athletes in stressful, evaluative contexts are more likely to make claims prior to performance like, "I am feeling tired," or "I am having personal concerns in this moment" (Coudeyville et al, 2008, p. 671; ; Martin & Brawley, 2002). Now, these claimed excuses may be wholly accurate or entirely fabricated, but the essence of an excuse lies in its function, not the truth value of its composite propositions.

As an example, Smith, Snyder & Handelsman (1982) assessed the claim self-handicapping behavior of individuals reporting high levels of generalized (dispositional) test anxiety. Participants first completed a survey measuring levels of text anxiety across contexts. Then they were led to believe that they would complete an examination which, the researchers (falsely) claimed, is a reliable measure of general intelligence and accurately predicts future success. This condition creates a direct link between the results and the respective identities of test-takers, likely motivating individual self-protective behavior.

Moreover, subjects were then randomly assigned to receive one of the following additional (again false) claims in addition to the test instructions:

(i) that test anxious individuals were likely to underperform[11] (excuse condition);

[11] "Although the [test] is a good measure of intelligence, it is negatively affected by the person's level of anxiety. In other words, people who are typically anxious or nervous when they take tests tend to get a score that is significantly below their true level of ability. These people…have better intellectual abilities than their score on the test would suggest" (Smith, Snyder, Handelsman, 1982, p. 317).

(ii) test anxiety has no effect on performance (no excuse condition)[12]; or,
(iii) no information about this relation (neutral).
In all, the test was divided into two main sections, and an anxiety questionnaire measured reported levels experienced in completion of part one; also included was a survey related to the effort exerted on the first portion.

When compared to individuals reporting relatively lower levels of generalized test anxiety, those high in trait test anxiety were more likely to claim situational anxiety prior to performance, which is exactly what we should expect; the nature of the assessment is inherently anxiety inducing and these people are most responsive to the relevant threat. However, when we look to intragroup differences, the findings become a bit more complex.

Those in the excuse condition reliably self-handicapped at a greater rate than those in the no excuse condition, irrespective of individual differences in general anxiety. The best explanation for this is that at least some individuals in the study took advantage of the explicit opportunity to self-handicap provided by the excuse, while perhaps others may have felt more empowered to voice genuine anxiety after being informed of its excuse value. Even though the evaluation is by its nature relatively nerve-wracking for test-takers, there is no reason for those low in dispositional test anxiety to report higher levels of state anxiety uniquely in those cases in which anxiety is expressly offered an excuse, unless this situational function of anxiety explains (at least, to some degree) why it is these individuals reported their state as such. Furthermore, since we see a rise in reported test anxiety despite levels of generalized anxiety, there is reason to believe that even those who sincerely experienced anxiety in the moment were more likely to report their state if made aware of its excuse value; this is likely responsible for some reports

[12] "One of the advantages of [this test], as compared to other intelligence tests, is that it is not in any way affected by other factors, such as test anxiety, which can make accurate measurement of intelligence difficult" (p. 317).

among those with greater levels of generalized anxiety. So, at least, it seems that the direct indication of anxiety as an excuse value motivates some people to report it, regardless of their actual affective state.

However, perhaps surprisingly, while the results showed that the explicit mention of the excuse reliably produced an increase in claimed state anxiety, the influence was not as significant as we might expect. Moreover, among high trait anxiety individuals, those in the excuse condition claimed test anxiety with *less* frequency than those provided no information. Therefore, those most likely to claim test anxiety were individuals with high levels of general anxiety who were provided no explicit information about the effect of test anxiety on performance. These individuals claimed test anxiety at a substantially higher rate than those in the no excuse condition.

Now, initially we may suspect that these findings expose a flaw in the excuse interpretation of self-handicapping. If people want to make an excuse for themselves, why not take advantage when the opportunity is right under our nose? The explanation is relatively simple: the most effective excuses are those in which the target does not recognize S's intention to excuse—exculpate—themselves from blame. Excuses work when the explanation is ostensibly true and the target does not suspect the excuse-maker to be engaged in an attempt to merely save face or *get off the hook*. When we intentionally offer a deceptive excuse to another, our efforts will fail if it is obvious that the expression is designed with the function of deflecting blame. Therefore, the findings suggest that high test anxious subjects felt the excuse to be "too obvious" for the audience when made explicit (Self, 1990, p. 54), but when the excuse is available without being rendered obvious, the findings suggest such a situation produces the optimal time to offer an excuse. Those in the excuse condition are aware of the potential function of self-reported anxiety, and perhaps some of those genuinely experiencing anxiety were positively motivated *not* to report it exactly because they did not want to make an excuse for themselves, or to be perceived as attempting to.

Indeed, self-reported test anxiety was most common among those with a high level of general test anxiety when the handicap was available, but remained implicit (Smith, Snyder, & Handelsman, 1982, p. 319). This much suggests that some strategy is involved here; individuals' reports are responsive to available self-protective resources. For more evidence of selective self-reporting, the analysis also showed that those in the no-excuse condition systematically reported lower levels of effort, with high trait anxious individuals reporting the lowest levels of effort. There is reason to believe, then, that the no-excuse condition motivates a search for another potentially available excuse across the board. Specifically, that high trait anxious people report the lowest levels suggests that those most threatened by the test are those most driven to find an exculpatory source of blame.

Undoubtedly, for some people in this setting and other similar evaluative contexts, the intention which underlies excuse-making is to deceive an external audience, or those who make the final evaluation; the intent is to mislead the audience into attributing test anxiety to the performance on the exam, all in hopes of manipulating casual attributions. For others, the claimed anxiety was likewise undoubtedly genuine, and an innocent representation of their inner experience, such that anxiety in fact was responsible for the outcome.

This suggests that self-handicapping behavior can be strategically employed, depending on the psychological set of the individual and the setting in which the task takes place. But, this much does not indicate intentional, nor even strategic, self-deception. The strategic element of self-handicapping is necessary to show intentional self-deception, but plainly insufficient. So, in turn, we need more evidence which points to intention in action in order to substantiate the claims related to self-deception. As we find, such evidence is particularly pronounced in instances in which the excuse-making venture manifests behaviorally, rather than as the mere claims of an individual.

Behavioral Self-Handicapping

Behavioral self-handicapping keeps much of the same structure and function as its verbal counterpart. Just like claimed self-handicapping, the action or pattern of actions are proactively employed as a means of facilitating self-serving attributions, except in these cases S *actively constructs or acquires* an impediment to success in the task prior to performance. In these cases, then, the excuse really does feature in an objective explanation of the outcome; S makes the task more challenging than required of them, and therefore these behaviors become an important determinate of the outcome. In case of failure, the obstacle deflects blame away from S, and should S succeed, however, the obstacle functions to make the performance look all the more impressive.

In their seminal study on behavioral self-handicapping, Jones & Berglas, (1978) demonstrated that participants could be manipulated into selectively consuming what they believed to be a performance-inhibiting drug when doing so provided a functional excuse for poor performance. Here, researchers instructed students that their participation was part of a trial comparing the relative effects of two new drugs on cognition functioning. In order to examine these effects, students were led to believe that each would complete two exams sequentially, and as usual these exams purported to measure innate intelligence. Participants were subject to one of the following two experimental conditions: In one, participants were given a set of questions containing four very easy questions and sixteen questions which contained no logically possible solution – *trick* questions, in other words. However, individuals in this condition nonetheless universally received feedback indicating success, no matter their actual scores. This is the "non-contingent" condition, and is a commonly employed set-up in experiments like these. The situation commonly creates an experience of difficulty and uncertainty in participants, since the insoluble questions should be perceived as particularly more difficult, and not immediately reconcilable with feedback indicating success. In the second condition, however, participants were subject to more mundane conditions, completing an exam which contained reasonably challenging, though ultimately solvable, problems. Further, in this condition, participants were given accurate feedback indicating success or failure – the "contingent" condition, since feedback corresponds to actual performance.[13]

The groups completed the first portion, at which point participants were given the opportunity to choose freely between one of two available drugs (placebos). It was explained that one of these drugs is ostensibly performance-enhancing while the other was described as producing performance-inhibiting effects, so subjects were led to believe that the subsequent exam would measure the drug's effect on performance but subjects were given the opportunity to choose freely.

In all, those in the non-contingent setting were significantly more likely to select the performance-inhibiting drug prior to completing the second round of evaluation, even after having been informed of its effects.[14] When researchers asked those in this group to justify their

[13] The average scores were between 12 and 14 correct answers out of 20 questions. With an average between 60%-70%, it seems that these problem sets were fairly challenging, but neither extremely easy nor difficult.

[14] Researchers found differences between the tendencies of male and female participants in this condition. 70% of males selected the performance-inhibiting drug, while only 46% of women chose this option. The role of sex differences in self-handicapping tendencies is a contentious issue, not particularly relevant to this work.

choice, recipients of non-contingent feedback often claimed a motivation to assist in or contribute to the study. That is, a substantial portion those in the non-contingent group claimed that “they were helping to make the experiment ‘work’ because, having scored 16 out of 20 on the first test, they had more room for downward than upward movement. Thus, [they claimed] the effects of the drug could be easier to see…” (p. 412).

There are a few things to notice going on here: First, this appears to be an example of some rather sophisticated rationalizing on part of participants in explaining the motivations for their decision against the alternative, which we will find to be a common theme in self-handicapping. Moreover, we should take notice that, while on the one hand, self-handicapping of this sort shields against future blame, on the other, it functions in protecting or enhancing an existing self-image by indicating a novel or unusual feature of situations producing failure which is likely absent in past cases of (non-self-handicapping) success. That is to say, self-handicappers may comfort themselves by concluding from the immediate evidence, “You know, it *is* true that the only time I fail at this kind of thing is when I do *X* [whatever the self-handicapping behavior]. I succeed when I give my best effort.” Even if this correlation between effort and outcome generally holds, however, the mere fact that S did not actively sabotage success in previous task iterations does not in turn entail that the success reveals any innate talent, though S will almost certainly interpret the evidence in this way. S’s self-belief as dispositionally talented, while perhaps not necessarily false, nor even immediately contradictory to the available evidence, is certainly not the belief best supported by the task outcome in the relevant context.

In all, the non-contingent condition induces unease in the participants about their levels of competency, as feedback indicating success likely contrasts with their internal experience and expectations given insolvable nature of the task set, while reinforcing self-doubt in the case of failure. Given that the task in this condition was composed primarily of *trick* problems, there is little reason to expect that participants experienced much, if any, internal confidence in their responses. However, after being presented with success feedback, those in the non-contingent condition ought to be motivated to accept positive feedback, and therefore subsequently self-handicap as a method of protecting and enhancing this welcomed self-belief. The Jones & Berglas study exemplifies the paradigm set-up for empirical investigations of self-handicapping.

The use of drugs or alcohol is only one common example of self-handicapping, and strategies may exhibit more or less nuance and finesse in locating, acquiring, or constructing a workable excuse. Generally, however, I will advance the claim that many instances of this behavior are instances of intentional self-deception. My arguments are substantially informed by the empirical literature, and yet even though this information is foundational to my arguments, there is simply a profusion of available evidence tracking the finer details of the motivations, strategies, and consequences associated with behavioral self-handicapping. Therefore, I will remain as faithful as possible to the empirical evidence, but I make no attempts to provide a comprehensive account of self-handicapping. I will do my best then to provide the most straightforward, *garden variety* cases of self-handicapping as situated in the discussion of intentional self-deception. The behaviors I have in mind ought to be familiar to most; my arguments will not rely on an especially involved understanding of the inner workings of self-handicapping. Proceeding, then, I provide some further empirical support that self-handicapping is *prima facie* self-deception, afterwards arguing for the Intentionalist account.

Section Four: Self-Handicapping Behavior as Intentional Self-Deception

In order to show self-handicapping as a *prima facie* example of self-deceptive cognition, I rely primarily on empirical evidence to bolster, what I hope, are common sense examples. There are plenty of brow-raising studies which illustrate a shrewd, wily self-handicapper who is capable of mending behaviors as demands require without missing a step. While interesting as these examples in fact are, the results do not supply the clearest representation of self-deception. So, after these perfunctory remarks, in my subsequent arguments about self-deception I employ a more familiar, standardly philosophical methodology.

The studies mentioned so far supply some elementary evidence that self-handicapping is self-deceptive in some cases. Most basically, recall the findings in (Smith, Snyder, and Handelsman, 198), which showed an increased propensity to self-handicap when the handicap remained available, but only implicitly – neither rejected nor explicitly expressed as an excuse. One interpretation of these findings is that, when expressly mentioned, the excuse is too obvious to be of use in manipulating an external audience's attributions. However, an alternative explanation is that perhaps self-handicapping behavior is most likely when the excuse is implicitly acknowledged, *because* the excuse provides an opportunity to self-deceive. On this understanding, implicit excuses are valuable because absent explicit acknowledgment of the excuse, not only does is the explanation more effective when directed externally, but the task context is moreover one in which S may self-handicap without reflexively acknowledging their own motivations and the function of their behavior.

Now, while the empirical literature produced in social psychology has consistently recognized the connection between self-deception and self-handicapping, though often directing analyses toward only the first of these potential explanations. For instance, Higgins & Snyder

(1988) define self-deception as “the process of having two antithetical self-relevant beliefs such that the more negative belief is motivationally held less within awareness” (p. 240), arguing that self-deception allows us to “deal with the fact that [we] have engaged in…excuse-making. The typical solution to this problem is for people to assert to themselves (and often to others) that they have…reasons for their seeming bad actions (i.e., making excuses)” (p. 241). Similarly Snyder and Higgins (1990) argue that as a feature of self-handicapping behavior the “primary importance of self-deception…[is that] it enables the individual to maintain his or her focus on external task concerns rather than being caught up in conscious self-contemplation regarding a negative performance (including the excuse, itself)” (p. 103). Said otherwise, the authors claim the greatest benefit of self-deception to self-handicapping is that it allows us to engage in the excuse-making behavior without being burdened by the conscious awareness of our own motivations or the function of our actions. This is surely correct as I see it, but the very fact that we must psychologically “deal with” our excuse-making behavior tacitly indicates that there is a role for self-deception in self-handicapping logically prior to that of suppressing awareness, and I am therefore doubtful that the *primary* importance of self-deception is found in this action-sustaining function.

Instead, this supervening function of self-handicapping behavior is that of tangible, self-deceptive evidence manipulation; self-handicapping *produces* evidential contexts conducive to self-deception. When the target audience for the excuse is internal (i.e., self-directed), self-handicapping behavior is that pattern of actions by which S manipulates the available evidence. So, if this function is primary, it follows that the value of self-deception as that state which sustains the behavior is only derivatively valuable insofar as it is, in turn, a means of achieving the broader, deceptive end of evidence manipulation. It is undoubtedly true that self-deception facilitates self-handicapping as initially explained, via the suppression of conscious awareness, but in taking a second glance at the relationship, notice that in positing such a role for self-deception, we only push the goalpost further back: if self-deception functions to suppress reflexive awareness of the underlying motivation or intention in excuse-making, what features of self-handicapping qua evidence manipulation necessitate this function (or nonetheless make it a useful one)? Though it may seem obvious, some brief analysis is justifiable for clarity's sake. So, why think that successful self-handicapping involves self-deceptive repression of our awareness as involved in excuse-making?

In arguing that the primary role of self-deception in self-handicapping is to repress reflexive awareness of the behavior itself, we risk overlooking the fact that self-handicapping is not itself an end which, by definition, is only achievable in virtue of this conscious *un*awareness. That is, there is nothing conceptually inherent to self-handicapping, as may be the case in self-deception, which renders concurrent self-awareness fundamentally paradoxical in the implementation of that behavior, and there is no *a priori* reason to believe that reflexive conscious awareness of self-handicapping is alone sufficient to render the plan ineffective. Instead, it must be something about the goal of evidence manipulation which renders self-handicapping impossible upon reflexive awareness of one's behavior. So, in order to understand the features of self-handicapping which render it potentially self-undermining, it is necessary to first indicate the primary goal of self-handicapping.

As we will find from the empirical evidence, the self-protective function of self-handicapping is usually fulfilled by S's retaining or enhancing a desired self-image. In order to achieve this, self-handicapping constitutes the behavioral process evidence manipulation which facilitates the self-deceptive maintenance of that image, and by definition the behavior works toward this goal at the expense of accuracy and objectivity. So, let's back up: self-handicapping is a means to a primary goal of self-protection. Therefore, it is this self-protective end which ultimately governs self-handicapping. If this self-protection requires of S the so described suppression of reflexive awareness, it must be that this self-protective is therefore the source of value for self-deception.

Though not accounted for within the self-handicapping paradigm elsewhere in the prominent literature, this relationship between self-awareness of successful self-deception is standard fare for the philosophy of self-deception as essentially a reformulation of the dynamic paradox. In these cases, self-handicapping just is a method of self-deceptive manipulation, and this in turn explains the self-undermining nature of reflexive awareness. Self-handicapping is employed toward the end of self-deception, and if we are to catch ourselves aware of our self-handicapping, this too implicates the potentially poor foundation of the beliefs predicated on self-handicapping. One cannot, for instance, self-handicap as a means of retaining a desirable self-image as a superior basketball player, while at the same time consciously aware of the behavior as designed to make ambivalent, or further corrupt, the proper causal attribution of my skill to the outcome of my play. So, when self-handicapping is geared toward self-protection, the behavior is the active process of self-deceptive evidence manipulation.

To see how this process comes together, we may look to the findings of Hirt and McCrea (2001) which showed that individuals who self-handicapped (in that case, failed to adequately prepare prior to the exam) performed worse on the relevant assessment than the average participant, but moreover, consistently attributed their performance to the handicap (failing to practice), and most interestingly, rated their competence in academic psychology (the subject of the assessment) as superior despite the significantly poorer performances on average (p. 1386).

Now, it is commonly recognized that self-handicapping positively impacts state self-esteem, as captured classification of the behavior as essentially self-protective, though the explanation for these findings is still somewhat contentious. In this study, however, researchers found that self-handicappers reported inflated self-beliefs about *specific domain* competence, and that this in turn mediated boosts to self-esteem. Compared to the non-handicapping group, self-handicappers consistently reported inflated self-beliefs about their competence in psychology, but interestingly, without reporting either an inflated sense of general academic ability nor competence in domains outside of academics. Moreover, high levels of self-esteem did not influence self-reported ability judgments; those with high self-esteem were not more likely to report superior ability in psychology. Rather, self-handicapping produced high levels of self-perceived domain competence among self-handicappers, in turn generating higher levels of post-test self-esteem.

Self-handicapping functions to retain or enhance ability beliefs about a fairly specific domain, which, in turn contributes to global self-worth (p. 1389). The mechanics of this process rely on the generation and content of *counterfactual* thoughts.[15] Counterfactual thoughts are just those which identify some way in which a past outcome might have turned out differently, and these play a crucial role in task performance and affect regulation. There are two kinds of these: *upward* and *downward* counterfactual thoughts. Upward thoughts indicate how a past performance might have produced a superior outcome ("If I would have studied more, I would have gotten a better grade."), and downward thoughts focus on the ways in which an outcome might have been worse than it was in fact ("At least I got a 'C'; I could have failed altogether."). Now, in typical conditions, upward thoughts induce negative affect by making salient some

[15] For a thorough review, see (Epstude & Roese, 2008).

relative shortcomings or disadvantages compared to an imagined ideal. Likewise, downward thoughts generally produce positive affect, since these highlight the ways that we are better off than we may have otherwise been. Interestingly, though, in the case of self-handicapping, the affective impact inverts, such that upward counterfactual thoughts engender a positive state, and vice versa.

For example, in McCrea (2008) the results showed that upward counterfactual thinking reliably produces positive affect, but only if the content of that upward counterfactual made salient an available handicap as an excuse. So, in this case, participants entered expecting to complete an evaluative assessment, and were then grouped in two: one group receiving the option of up to ten minutes of study time prior to the assessment (practice condition), while the other group received no such chance (no practice condition). In order to induce handicapping via a failure in preparation, researchers claimed to those in the no practice condition that a computer glitch mistakenly produced uneven groups, and that too many people were assigned to this condition. Then, these participants were asked if they would not mind joining the practice condition for the sake of the experiment. Likewise, in the practice condition, participants were led to believe that the no practice condition had been over-assigned, and subsequently asked to the other condition, ostensibly forgoing practice for the sake of even group sizes and fair experimental results. Researchers, however, reiterated that the decision was the participant's to make freely, and that there would be no repercussions for declining to switch conditions. Importantly, all participants received non-contingent feedback of failure, so the motive to self-protect ought to have been present and salient.

Finally, after completing the exam, participants were asked to report any of their thoughts structured as, "If only I… / At least I…", thus priming responses in the form of upward and downward thoughts. Respondents also reported their current mood, state self-esteem, and most importantly, the degree to which they attributed their effort in practice to their performance. Those in the practice condition generated significantly fewer upward thoughts related to practice, while those in the no-practice condition regularly mentioned practice time in their responses. Participants believed that they had freely chosen whether or not to practice, so for those in the practice condition, it follows from a self-protective motive that this group ought to generate upward thoughts unrelated to practice, in order to deflect responsibility from themselves. By that some token, participants who chose to switch conditions and forgo practice reliably appealed to this lack of preparation in reporting upward thoughts. So, somewhat paradoxically, those who *chose* to practice were less likely to indicate that practice as responsible for the outcome when feedback revealed failure, while those who actively turned down the opportunity to practice readily suggested this lack as the explanation for failure. Whether practice was a salient feature in the content of counterfactual thoughts depended upon whether the practice condition exculpated the participant of their performance; in the case of failure, choosing not to practice explains away the outcome, while failing in spite of practice deals an even greater blow to the self-image. The empirical results from this study are congruent with this explanation.

By making salient the lack of practice in explaining task failure, upward counterfactuals afford the opportunity for S to locate the source of blame externally; at least in the sense of external to S's inherent or innate features. Apparently, even though these individuals believed that they had chosen to skip practice, this lack of preparation accounted for the contents of their We have good reason to believe, then, that not only were these individuals motivated to blame inadequate practice time for failure, but that these attributions occurred without accompanying contrition related to the participant's (perceived) free decision to skip preparation. In short, these participants were ready to blame a lack of practice, but did not experience guilt in failing to prepare when given the chance.

Interestingly, though, given the self-inflation consequent to these upward thoughts, those in the no-practice condition, were less likely to report an intention to practice in the future. That is, even though the self-reported upward thoughts of those in the no practice condition regularly indicated a lack of preparation as responsible for failure, these participants reported no resultant increase motivation to practice in the future. However, researchers then altered the conditions as to lead participants to believe that each individual would receive two attempts at the assessment; and in this case, upward counterfactuals in the practice condition then predicted an increased intention to practice in the future. When given only one opportunity for evaluation, it seems that those in the no practice condition were satisfied to reflect upon an excuse, but were not motivated by this reflection to alter future behavior, suggesting that the individuals were motivated by self-protection at least slightly more than performance or accurate diagnostic information.

Thus, these findings suggest that counterfactual thinking plays a role in forming intentions, and thus in future task performance, only if accompanied by negative affect – only if the upward thought does not make salient a handicap which excuses the performance. And, upward thoughts which enhance or maintain affect do not reliably motivate a future intention to exert more effort in preparation for the task, suggesting the self-protective reward is the sought after benefit, even if sacrificing domain competence as a result. Supporting these findings, Flamm and McCrea (2012) found that when people are motivated by "improvement or have already addressed self-protection concerns, upward counterfactuals concerning a lack of effort increase subsequent persistence and facilitate performance….On the other hand, when individuals are faced with an ego-threatening failure, they may generate and interpret upward counterfactuals in a manner that excuses this poor performance…By reducing self-blame and negative affect, excusing upward counterfactuals undermine motivation to improve in the future." (p. 379).

Similarly, Mercier (2017) examined the role of "prefactual" thinking in self-handicapping, and in relation to counterfactual thoughts. Prefactuals, contra counterfactuals, are just those thoughts which indicate how outcomes might be different *in a future* iteration. In this study, participants received random feedback indicating success or failure[16], and were then primed[17] to generate either prefactual or counterfactual thoughts about the outcome. Results showed that the responses in the counterfactual condition, irrespective of success or failure feedback, tended to make salient uncontrollable features of the test situation. These features were sometimes internal to the person (i.e., "If I were good at this sort of game"), while others focused on external factors ("If more time was allotted for the test").[18] Yet, although this pattern held in both the success and failure condition, the propensity to offer counterfactuals was most pronounced in the case of failure, which Mercier and his colleagues argue indicates a tendency to (quasi-)spontaneously generate exculpatory counterfactual thoughts in the event of failure.

In another iteration of Mercier (2017), researchers tasked participants with a timed logic assessment, only this time providing immediate, contingent (accurate) feedback. In the task,

[16] Compared to an imaginary average.

[17] So, those in the success condition were instructed to recall how things might turn out worse next, and those in the failure condition were instructed to recall how things might be better next time. To 'prime' a counterfactual response, participants were instructed to complete the phrase, "Things would have been better for me if…" Alternatively, to prime prefactual responses, researchers altered the phrase to read, "Things will be better for me in the next game if…" (p. 263).

[18] There is no data provided indicating the frequency of internal or external explanations.

participants we given a set of 10 logic problems and instructed to complete as many as possible within a twenty second limit; completion of all ten problems within the timeframe indicates success, and completion of any less than 10 in this time was deemed a failure. Here, participants were again randomly assigned to a counterfactual or prefactual condition, rather than allowed to spontaneously and independently generate thoughts. The experiment replicated findings from the previous study indicating a general propensity to generate uncontrollable counterfactual thoughts, though perhaps indicating a more pronounced motivation to self-protect when presented with accurate feedback. Irrespective of success or failure, only 12% of counterfactual responses appealed to controllable behaviors in this study (p. 265).

Fleshing out the distinction, the content of controllable counterfactuals makes salient the aspects responsible for failure which are seemingly under out control, and are structured roughly as “If I had *done* X [instead of Y] things likely would have turned out better.” So, it makes sense that those disposed to self-protect are least likely to generate these thoughts; the content of this thought is strictly antithetical to the self-protective project since controllable behavior is that which is under the individual’s control, that which could have been done otherwise, thus directly place blame in the hands of the individual in question. Alternatively, in the prefactual condition, researchers led participants to believe that the study included two rounds of assessment, and respondents reported their thoughts in between sections. Likewise, when primed to generate prefactuals, the content of the thoughts highlighted controllable features of the task performance. So far, then, results indicate a propensity to focus on uncontrollable aspects of the task situation when recalling past failures, and the opposite tendency to generate thoughts related to controllable aspects of our own behavior when framed in light of a future goal.

Most interestingly, though, researchers finally altered experimental conditions once more, this time providing accurate feedback, but furthe instructing some participants to offer responses which ought to be valuable to a "future self" or another test-taker who would complete the assessment at a later time (p. 265). Therefore, here the researchers deliberately requested a prefactual response meant to focus on controllable aspects of the performance. When explicitly instructed to provide this kind of advice, participants from both conditions reliably generated prefactual thoughts that focused on controllable aspects of the task situation. Together, these findings suggest a general preference for uncontrollable counterfactuals in virtue of their self-protective use, but also that participants nonetheless retain some objective, diagnostic information that may be introspectively accessed with relative ease.

This diagnostic information is that which an individual ought to be most motivated to avoid in self-protection, since it highlights our own potential responsibility for failure and the ways we might improve upon our performance, both of which locate the individual as the primary source of responsibility. The only apparent variable responsible for this shift in perspective is the introduction of the prefactual (advice) instruction, therefore suggesting that a shift in goal-orientation is sufficient for the generation of objective, accurate responses. Participants compelled to abdicate themselves of responsibility, therefore, are seemingly not ignorant of the proper attribution, but apparently not motivated to attend to that evidence.

The Nature of Belief in Self-Deceptive Self-Handicapping

From the findings presented, we have good reason to believe self-handicapping behavior is intimately related to, or reliant upon, self-deception in many cases. Hirt and McCrea (2001) suggest that self-handicapping functions in the retention of valued self-beliefs related to fairly specific ability, in turn contributing to global self-esteem as a result. This pursuit of a desired belief is a paradigm instance of the motivations which drive self-deception. Furthermore, McCrea (2008) argues that we receive an esteem boost when generating excuses for our failures, so long as those excuses do not implicate a valued feature of our identity. And, in these situations, we saw that such reflection impedes future intentions to practice, suggesting a stronger motivation to self-protect by means of self-belief regulation rather than by way of skill development or task performance.

Pulling it together, some of the strongest evidence of self-deception is from the Mercier (2017) study, which indicate that a rather thinly veiled wall shields objective information from interfering with the maintenance of a desired self-image. As we saw, when not primed to frame our thoughts as future-directed, our default thought processes focus on uncontrollable features of previous task situations in which we previously failed. But, even those who initially offer a counterfactual explanation were capable, with apparently little effort, of producing useful advice the content of which made salient controllable behaviors, and seem instead to avoid such thoughts unless prompted by deliberate request or by a situational reframing of the goal (Flamm & McCrea, 2012). In conjunction, these findings suggest that objective information is available to the self-handicapper, but that these individuals may selectively repress such knowledge toward the self-deceptive end.

Now, we may turn back our analysis to the original explanation of the of self-deception in self-handicapping. Snyder & Higgins (1990) are right to point out that a key function of self-deception in sustaining self-handicapping is that “by being unaware of our excuses, we avoid having to excuse ourselves for making them” (p. 103). We have seen that self-handicapping reliably produces enhanced self-beliefs and boosts to self-esteem, and unless the excuse making intention is suppressed from consciousness, it is unclear how we might account for the measurable, positive impacts on the self. The fact that individuals reliably rate themselves as superior in the relevant domain after successfully self-handicapping, and in turn experience gains in mood and global self-esteem, indicates more that these folks believe, rather than simply *avow*, inflated self-beliefs related to domain skill.

In all, we may begin to notice a two-tiered, or an internal/external, structure in the relation between self-deception and self-handicapping in these cases. That is to say, the self-deception which facilitates the self-handicapping process itself via suppression of conscious awareness is subsidiary to the self-handicapping behaviors since these are means of self-deceptive evidence manipulation. The process is futile without the second-level deception. The second-level deception is fundamentally internal, by which I mean that the deception is constituted primarily by the selective rationalization and retrieval of values, memories, or motivations, and not directed at corrupting the external evidence. The self-handicapping behavior is then external self-deception, in that this process involves actively altering or blurring the available evidence. Nonetheless, however we might best classify the self-deceptive cognitions in self-handicapping, no robust categories are required in advancing my claims that intentional self-deception is exemplified in the behavior.

What seems clear, though, is that this process involves something like the psychological partitioning accounts espoused in Davidson (2004) and Pears (1984). Recall, these accounts claim that the mind is segregated, "partitioned," and that beliefs may mutually reside in distinct realms without conscious contradiction. The empirical evidence suggests a great deal of self-handicapping involves the insulation of objective diagnostic information, motivations in task approach, and values or desires, from the self-handicapper's conscious explanations for performance. However, as we saw, shifts in motivation often entail a revision in the self-handicapper's epistemic framework toward a more objective, less vicious orientation. The suppression of undesirable conscious awareness is therefore best explained as a strategy in pursuit of self-deception, since it is the transition away from the goal of self-protective belief formation which apparently halts the self-handicapping process, and thereby renders such repression obsolete.

In all, we see that self-deception facilitates self-handicapping in masking the individual from their own motivations, while the self-handicapping behavior itself is self-deceptive in the active sense, as the method by which one deceives themselves. In all, it is likely unhelpful to chart this process in terms of which self-deception is temporally prior to the other, since the behavior and cognitions are mutually reinforcing, and there is not evident method by which one might mark the end of self-deceptive thought suppression compared to the beginning of behavioral self-handicapping. Likewise, behavioral self-handicapping undoubtedly affects not only domain specific self-beliefs, but our behaviors also alter our propensity and ability to engage in the second-order, self-deceptive suppression of motivation. The process is cyclical.

Self-Handicapping & Intention

Imagine Alex, a struggling law student who is threatened by the prospect of failing her final exams. She came to the program thinking of herself as naturally superior to others in terms of intellect, having been handed success for most of her life. However, she is now beginning to wonder about the accuracy of her self-image, as the work in law school is significantly more challenging than that to which she is accustomed. She worries that if she fails to prove her intelligence on her exams, then her entire self-image will come into question. Not only is this outcome threatening to Alex's global well-being, additionally she worries whether she is even capable of burdening such an identity crisis while maneuvering her schoolwork. Though, for all these creeping doubts, Alex nonetheless believes in her innate advantage over others, and is deeply motivated to retain that self-image.

As final exams approach, Alex cannot seem to garner the motivation to study, finding herself catching up on lost sleep or watching her favorite television programs for hours on end. Alex knows that prudence dictates studying as the most reasonable course of action, and feels guilty for her lack of motivation. Ultimately, however, Alex reassures herself that her talent is sufficient to overcome any deficit in preparation. To build confidence, she recalls all of the times in her college calculus course that she aced the exams without studying at all, and these bring some relief to her guilt. Finally, on the night before finals, Alex tries to cram as much reading in as possible, scrambling to make up for lost time.

She arrives to the exam sleepy and in no shape to perform her best; still, she reflects on all that success in calculus to quell her worries. Alex exits the exam, though, feeling quite confident about her performance, and relieved to finally shed her concerns about performance and preparation. And, when grades are returned, Alex finds that she her results are congruent with the class average, finishing right around the 50th percentile. Not a terrible outcome, but not that to which Alex is familiar, and surely not one indicative of innate academic virtue. “Well, I know I could have done it if I really tried,” Alex thinks, “I know I am still smarter than the other students. I am sure I will show them next time.”

Alex is self-deceived. She is using her lack of preparation to excuse an average grade, and in doing so retains her self-belief as intellectually superior. Whether the long-term evidence supports her self-image as intelligent is a separate question from the immediate deception: the evidence available, her preparation and performance, warrants an epistemic *down-shift* in the strength of Alex’s self-beliefs, but instead this evidence functions to strengthen these beliefs. If

Alex were to engage in a good-faith effort in studying and otherwise preparing for the exam, but only to come up short in the grade, she must confront unwelcomed evidence which suggests that, even at her best, her self-image as natively intelligent is fatally gilded. But, by proactively engaging in behaviors which clearly facilitate underperformance, Alex creates a context in which an underperformance can be attributed to those behaviors. Alex infers a belief about her own intellectual superiority and innate qualities from evidence which supports a conflicting, if not contradictory, conclusion.

Mele (2001) mentions examples which, on the surface, appear quite analogous to those behaviors described thus far:

> "Ann believes that she can cultivate the trait of kindness in herself by acting as if she were kind; so, because she wants to become kind, she decides to embark on a program of acting as if she were kind, and she acts accordingly. Because Bob would like to be a generous person, he finds pleasure in actions of his that are associated with the trait; consequently, Bob has hedonic motivation to act as if he were generous, and he sometimes acts accordingly. Unlike Ann, Bob is not trying to inculcate the desired trait in himself.
>
> … It is easy to imagine that, after some time, Ann and Bob infer, largely from their relevant behavior, that they have the desired trait, even though they in fact lack it. However, from the facts that these agents want it to be true that *p*, intentionally act as if *p* owing significantly to their wanting *p* to be true, and come to believe that *p* largely as a consequence of that intentional behavior, it does not follow that they were trying to deceive themselves into believing that *p* or trying to make it easier for themselves to believe that *p*. Ann may simply have been trying to make herself kind and Bob may merely have been seeking the pleasure that acts associated with generosity give him." (pp. 19-20)

Ann's case can be excluded from the outset, since her motivation is to "inculcate the desired trait" in herself, while the motivation to self-handicap is not to develop or embody a trait, but to retain the self-belief about the possession of that trait. The case of Bob, who merely desires to act generously, without any motivation to be a generous person, is a bit of a trickier case. Suppose Bob is placed in some situation in which levels of generosity are salient, but for whatever reason, Bob is particularly dissuaded from acting generously but retains his desire to be generous. Then, suppose he constructs some obstacle in order to self-handicap and form an excuse explaining why, on *this* occasion he cannot display his claimed or desired levels of generosity. Whether motivated by hedonistic desire, in both cases Bob may act *as if* he is generous, retaining his favorable self-image, but may accomplish this through very different behaviors.

What we see is that Bob may come to believe himself to be generous either by self-handicapping or by plainly acting in a generous fashion; and, as Mele notes, sometimes he does act as such. The fact that Bob engages in pretense via self-handicapping, as opposed to the genuine behavior, indicates the substitution of valued belief retention in place of generous behavior. Ann's and Bob's respective cases may illustrate examples of agents who are self-deceived on account of inferred explanations for their own behavior, and biased interpretations of these observations fuel self-deception here in much the same way as in self-handicapping. But, whatever Mele's original examples say about self-deception, they are not instances of self-handicapping.

Importantly, the presence of self-handicapping behavior is evidence of a self-deceptive motivation, and in asking why self-handicapping is only occasionally employed, we basically rehash the "selectivity" problem (Burmudez, 2001). The selectivity problem is levied against Motivationalist accounts on grounds that "the desire that p should be the case is insufficient to motivate cognitive bias in favour of the belief that p… How are we to distinguish these from situations in which our desires result in motivational bias?" (p. 317). For the self-handicapper, both a good-faith effort which leads to success and deceptive behavior might produce the same self-serving benefits, but obviously these behaviors are mutually exclusive. So, supposing the self-imagine concerns might motivate either behavior, it is not clear how the deflationary account explains the employment of self-handicapping in terms of pure motivation. The Intentionalist explanation is that the self-handicapper engages in the behavior because of an intention to self-deceive, in conjunction the fact that self-handicapping is just such an effective means of successfully implementing that intention. A closer inspection reveals this, I argue, to be intentional self-deception. However, before I can adequately explain my position here, it is beneficial in understanding the selectivity problem, and the Motivationalist response, to take a closer look at Mele's position.

As we have seen, according to Mele, many motivations may substitute for intention as that which explains biased evidence manipulation. So, accordingly, on the deflationary position, it is this motivation which not only sparks the self-deceptive goal, but also that which sustains it. For Mele, this motivation alone is responsible for self-deceptive bias, and we recall that such accounts argue that no intention is required to guide and monitor the biased judgment of evidence (2001, ch. 2-3). At the core of Mele's account is the notion that the self-deceiver is primarily engaged in the practice of avoiding costs incurred by the false acceptance or rejection of the given proposition, such that self-deception most often occurs when acceptance of the false belief that *p* is less than the costs associated with falsely rejecting the true proposition, not-*p*. However, as also mentioned, Mele allows for extreme, fringe instances of intentional self-deception. For example:

> "Ike, a forgetful prankster skilled at imitating others' handwriting, has intentionally deceived friends by secretly making false entries in their diaries. Ike has just decided to deceive himself by making a false entry in his own diary. Cognizant of his forgetfulness, he writes under today's date, 'I was particularly brilliant in class today,' counting on eventually forgetting that what he wrote is false. Weeks later, when reviewing his diary, Ike reads this sentence and acquires the belief that he was brilliant in class on the specified day." (p. 17).

So, while Mele concedes the theoretical coherence of intentional self-deception, he argues that such instances are anomalous. However, self-handicapping as illustrated is not anomalous, and moreover, it is not clear that Mele's example illustrates a case of intentional self-deception, because for Ike, whether he self-deceives is dependent upon whether he forgets his intention in writing the journal entry to himself. This forgetting is surely not intentional, and as such I would argue that Mele has not here shown an instance of intentional self-deception. Not only does Ike forget his intention, which seems to be more than a matter of volition, but at the time of belief acquisition, he seems to have lost the intention full stop.

While I do not wish to categorically deny the possibility of intentional forgetfulness, this example does not illustrate the necessary causal link between the initial intention and the resulting self-deception. Rather, Ike forms an intention to self-deceive, in turn predicting that his forgetfulness may help accomplish this end, but ultimately whether he forgets his primary goal is not a matter of intention. His intention to self-deceive does not *explain* the acquisition of false belief, but rather his forgetfulness does, and so there can be no self-deception in cases in which Ike fails to forget, which is ultimately a matter beyond his control. Thus, the link here between intention and deception is insufficient to call such self-deception intentional. In the same way that I may form an intention to win a hand of Texas Hold 'Em, I nonetheless cannot rightly say that I *intentionally* won the hand when the river card turns up in my favor. Even though I might have possessed the intent to win, it does not entail that I intentionally brought about those circumstances, since no matter the strength of efficacy of my intentions and practical reasoning, a key factor in determining the outcome is fundamentally out of my control. Well, this is like Ike. His self-deception is best explained by his forgetfulness, over which he does not possess control in a manner sufficient to render his forgetting intentional in nature. Ike’s case is not the strongest example of intentional self-deception, and Mele is right that, if we assume the phenomenon is explained as such, then we need pay it little attention. But self-handicapping offers a significantly better illustration.

Recall Alex, the self-deceived law student. She is faced with finals, and is motivated to protect her self-image. She could accomplish this by studying diligently and performing well, but this method risks the possibility of failure even after a good effort, which is the worst possible outcome for her self-image. Due to some doubt regarding her odds of success, however, the self-handicapper employs behaviors which they know are likely to impede success, and in doing so offering themselves a self-deceptive compromise. The desire to maintain a positive image of self may reasonably motivate the selection of either a good-faith or self-handicapping effort. This desire, then, cannot alone explain instances of self-handicapping behavior compared to good faith effort, since, keeping the desire constant, we have reason to believe that self-handicapping is adopted as a more effective means of goal pursuit.

This is not like Ike, because his motivation is only correlated with the outcome if, by chance, Ike forgets his goal, while in cases of intentional self-deception as I understand it, the relationship between intention and outcome is more intimately connected. However, this alone does not render the deflationary position inadequate, since one may claim that this same desire motivates individuals to avoid threatening hypotheses, pursue pleasant ones.

Upon closer inspection, though, this explanation will not suffice. What would such a motivation look like? We may account for internal biasing in these cases via an appeal to motivation alone, but the desire to believe that *p* in these cases is in fact antithetical to the self-handicapping behavior as a means of evidence manipulation. We can explain a case in which one fails at some task, and in turn rationalizes that failure in a *post hoc* fashion, but this explanation does not account for the behavior engineers or manipulates evidential circumstances. Internal biasing, then, does not the type of intentional self-deception as indicated in this work, but functions as a prerequisite or concurrent state that is necessary for S to sustain their self-deceptive project.

Self-Handicapping as Self-Deceptive Action

Recently, Marcus (2019) advances similar arguments, claiming that deflationary positions fail account for instances of "self-deceptive action," as "cases in which someone is not (at least in a straightforward sense) lying, but yet disavows a correct description of her intentional action" (p. 1205). The arguments, in short, is as follows:

> "To say that S is y-ing in order to x is to portray her as already understanding herself as engaged in the project of x-ing, and adopting the means y-ing for the reason that it serves what she already understands to be her end. The fact that she understands herself as x-ing is then what explains why…if she stops believing that her y-ing is a means of x-ing—then she will stop y-ing. The realization directly bears on what she already understands herself to be doing. She might then start z-ing in light of her instrumental belief in the connection between z-ing and x-ing…. Even in cases of self-deception, the agent persists in the relevant means-action only insofar as she takes the means action to be realizing the desired end, and therefore only in virtue of the fact that she understands herself as pursuing that end" (p. 1217).

Applied to self-handicapping then, the argument would entail that self-handicappers realize that they are self-handicapping as a method of self-deception, and this is borne out by the observation that such behavior answers the "how shall I question of self-deception.

Thus, we see the two-fold nature of self-deceptive self-handicapping. Like Alex, the self-handicapper behaves as such in order to retain a self-belief, which is one ultimately unsupported by the evidence. But, as a practice in self-deceptive action, the self-handicapper is deceived about their motivations in acting and the function underlying the behavior. Self-handicapping is intentional self-deception insofar as the behaviors are intentional actions, motivated by the goal of self-deceptive belief formation, and it this intention in concert with the subsequent behavior which explains why it is that S falsely believes that *p*. Note, I do not claim that self-handicappers intentionally deceive themselves as a matter of the second-order deception which sustains self-handicapping. But, this does not threaten my case; for, whether or not the self-deceptive repression of conscious motivation is intentional, the behaviors which constitute the evidence manipulation responsible for the inflated self-beliefs are intentional and pursuant to self-deception. Self-deceptive self-handicapping is the result of S's settling on self-deception as a means of self-protection, and the behaviors function to bring the self-protective belief about.

The Selectivity Problem

As previously mentioned, the "selectivity problem" (Bermudez, 1997; 2000) is a problem which potentially confronts both the deflationary and Intentinoalist position. It is often taken to be a virtue of Intentionalist accounts that these present a (ostensibly) straightforward answer, that self-deception occurs when the desire results in an *intention* to form (or to make it easier to form) the self-deceptive belief. However, as indicated in (Jurjako, 2013; Funkhouser, 2019), Intentionalist accounts likewise are burdened with explaining the reasons for which the intentions themselves arise in some cases and not in others. Just as desires alone are insufficient to produce false belief, such states are likewise alone insufficient to explain an intention to bring about some belief.

Mele responds to the selectivity problem by arguing that, since presumably there is a difference in the contexts in which self-deception occurs and those in which it fails, there is no evident reason to suggest that this contextual difference may not just as well explain the selective nature of self-deception without any further appeal to intent (2001, p. 62). In short, Mele requests, "I would like to be told how alleged intentions to bring it about that one believes that p are successfully executed in garden-variety cases of self-deception" (2019, p. 9). He claims this challenge is so far unmet, and while I cannot provide an answer which settles the matter, we may point to some potential progress. In self-handicapping, undoubtedly situational features play a key role in determining the formation of an intention to self-handicap. As we have seen, these factors include the relationship between task outcome and self-image, the type and difficulty of the task at hand, as well as the availability and salience of a potential handicap. At minimum, the intention to self-deceive may only be carried out in those contexts in which the excuse is available and utilizable. One may form the intention to self-deceive in various situations, but that intention may only be realized in contexts which provide the adequate resources.

The intention arises in the first place, however, as a matter of the perceived task context. Since self-handicapping is fundamentally self-protective, then we would predict the self-deceptive intention to arise in situations in which self-handicapping best serves self-protection. The Intentionalist account (or anyone else) cannot be expected to provide a full "etiology" (Mele, 2019, p. 7) of self-deception, and I do not make any such claims in this work. But, the contemporary movement in self-deception perhaps makes more of a mountain of this problem than is warranted. So, it certainly is puzzling to imagine how and why one may form and implement an intention to self-deceive in cases in which the intention is just to produce a self-deceptive state. But as argued in Bermudez (2001) the intention need only be to bring about *some* belief, not necessarily to bring it about that S is self-deceived, and this is the case at least in self-handicapping. In self-handicapping, these individuals are motivated to self-protect, which may be accomplished via material success or self-handicapping. Thus, the intention to self-deceive is formed on the basis of situation constraints which render self-handicapping a superior means of achieving the self-protective end. Thus, in all, the dynamic paradox is paradoxical in these instances only if we fail to consider intentional self-deception as aimed at something other than the self-protective end. S intentionally engages in self-deception as a means toward that broader goal, and qua implementation intention, there is nothing incoherent about intentional self-deception as such.

Does This Show Intention?

Now, it may be wondered, could we not still explain the behavior without an appeal to intention, given the concession that self-deception is context-dependent in this way? I am not sure how. As argued earlier, the avoidance of objective feedback and the sacrifice of material success does not seem to be explainable by pure motivation, and is seemingly antithetical to the desire that *p* (where *p* represents the relevant competency). What deflationary accounts must maintain in these cases is that the desire that *p* motivates the self-handicapper's biased interpretation of their behavior and its outcome in producing self-deception. But this does not explain the empirical evidence which indicates the level of strategy employed in self-handicapping, and more importantly cannot explain why these individuals do not engage in a good faith effort, instead self-deceiving in the face of potentially negative evaluations. The behavior is strong evidence that S is *settled* on adopting self-deceptive means in order to self-protect, since we know that agents may just as well explain away their failures retrospectively, but her opt for a proactive approach. The deflationary position does not explain why S chooses self-handicapping over some merely verbal excuse, and the fact that self-handicapping is anticipatory in nature strongly suggests foresight about the function of that behavior.

Beginning to wrap up, a particular empirical study that is unrelated to self-handicapping but intimately related to self-deceptive action is (Quattrone and Tversky, 1985), which showed that individuals will self-deceive about their intentions in acting if doing so provides some benefit. Specifically, these participants apparently self-deceived about their level of effort in keeping their arm submerged in ice water when led to believe that such cold tolerance was an indicator of heart health. However, while these individuals did submit themselves to the conditions for longer, they did not report an increased effort to do so. The explanation offered is that these individuals were motivated to increase their efforts in keeping their arm submerged without admitting to themselves their reasons for doing so. If aware that they intended to submerge their arms for as long as possible, there is no reason to believe the experiment evidences an actual increased tolerance, and likewise provides no evidence of heart health.

Mele claims that there is no need to posit intention here, since it may simply be the case that when presented with evidence of heart health, the motivation for this to be true biases understanding of their own efforts. Perhaps these individuals "hope" that they will find evidence of a healthy heart, and this desire in turn biases their interpretation of the available evidence (2019, p. 13). This is all likely correct, and may well explain the internal self-deception as the suppressed awareness of one's motivations and intentions. But it also overlooks the intentional self-deception as evidenced in external evidence manipulation. That being, these participants, like self-handicappers, settle on engaging in an effort in order to bring about favorable evidence, and *this* settling is the relevant intention. To use some admittedly loaded language, a lay explanation of the behaviors might describe the individuals as "deciding" to engage in the activities which make it easier to believe that *p*, and as such, form an intention to self-deceive and attempt to see that intention through. Though I provide no thorough explanation of how it is that individuals self-deceive about the content of their subjective reasons for acting by way of internal biasing, this does not alter our observation that in anticipating some outcome, individuals apparently form an intention to make it easier to acquire or retain a welcomed self-belief, and further that individuals are often successful in bringing those beliefs about.

This is the fundamental benefit of framing self-deceptive self-handicapping within self-deceptive action. As Marcus (2019) points out, what the deflationary position cannot evidently explain is the action guiding nature of the behavior without an appeal to intention, but what I hoped to have done here is supplement this argument, showing a case in which this action itself generates self-deceptive beliefs. The behavior is intentional in the banal sense of non-coerced, but is an instance of intentional self-deception on grounds that self-deception sparks, sustains, and satisfies the practical reasoning manifest by the intention.

Lastly, along such lines, the behavior may be explained as an instance of "robust, unconscious self-deception" (Funkhouser & Barrett, 2016), which is strategic and nuanced, but does not posit the presence of an intention in self-deception. That work, however, does not outright deny the possible function of intention, but rather wants to "side-step" the issue (p. 683). But, in all, it is unclear what evidence of intention might satisfy the staunch Motivationalist beyond this kind of strategic manipulation. Mele does not deny the possibility of unconscious intent, and describes intention as an executive attitude of S's being *settled* on a course of action, where this intention guides and sustains that course. In all, this sounds very much like the case of self-deceptive self-handicapping. The self-handicapper is disposed to self-protect, and adopts an unconscious attitude of being settled on self-deception as a means toward that end. Thus, in these cases, S forms the intent to self-deceive in light of the broader goal of self-protection, and self-handicapping operates as a means of carrying out the self-deceptive venture. If successful, self-handicapping ultimately terminates in the acquisition or retention of a valued, self-deceptive self-belief.

Conclusion

I have offered a *prima facie* case that self-handicapping behavior may constitute an instance of self-deception, and moreover, is best explained by Intentionalism. As argued, self-handicapping is motivated by, and results in, the same states as self-deception; specifically, a desire to believe some proposition. Furthermore, I have argued that the nature of self-handicapping as anticipatory, and as a self-deceptive *action*, renders the best explanation one which appeals to intention. Motivationalist accounts cannot readily explain the selection of self-handicapping over a good faith effort in situations which share a self-protective goal without an appeal to the intention to self-deceived. Admittedly, this does not bar a more sophisticated Motivationalist reading from explaining the phenomenon, but as a new investigation in the field of self-deception, no such literature exists. For now, it appears Intentionalism is best suited to account for the behaviors and beliefs of the self-deceptive self-handicapper.

References

Adams, F., & Mele, A. (1989). *The Role of Intention in Intentional Action*. Canadian Journal of Philosophy, 19(4), 511-531

Audi, R. (1982). *Self-Deception, Action, and Will*. Erkenntnis, 18(2), 133-158.

Bach, K. (1997). *Thinking and Believing in Self-Deception*. Behavioral and Brain Sciences, 20 (1), 105-105.

Baghramian, M. & Nicholson, A. (2013). *The Puzzle of Self-Deception*. Philosophy Compass, 8 (11), 1018-1029

Berglas, S., & Jones, E. E. (1978). *Drug Choice as a Self-Handicapping Strategy in Response to Noncontingent Success*. Journal of Personality and Social Psychology, 36(4), 405–417

Bermudez, J. (1997). *Defending Intentionalist Accounts of Self-Deception*. Behavioral and Brain Sciences, 20 (1), 107-108.

Bermúdez, J. (2000). *Self-Deception, Intentions and Contradictory Beliefs*. Analysis, 60 (4), 309-319.

Coudevylle, G., Martin Ginis, K. A., & Famose, J.-P. (2008). *Determinant of Self-Handicapping in Sport and Their Effects on Effort*. Social Behavior and Personality: An International Journal, 36(3), 391-398.

Davidson, D. (2004) *Problems of Rationality.* Oxford University Press

Epstude, K., Roese, N. *The Functional Theory of Counterfactuals*. Personality and Social Psychology Review: an Official Journal of the Society for Personality and Social Psychology, Inc, 12(2), 168–192.

Funkhouser, E. (2005). *Do the Self-Deceived Get What They Want?*. Pacific Philosophical Quarterly, 86 (3), 295-312.

Funkhouser, E. (2009). *Self-Deception and the Limits of Folk Psychology*. Social Theory and Practice, 35 (1), 1-13.

Funkhouser, E., Barrett, D. (2016). *Robust, Unconscious Self-Deception*. Philosophical Psychology, 29 (5), 682–696

Funkhouser, E (2019). *Self-Deception*. Routledge.

Higgins, R. L., & Berglas, S. (1990). *The Maintenance and Treatment of Self-Handicapping*: *From Risk Taking to Saving Face and Back.* In R. L. Higgins (Ed.), "The Plenum series in social/clinical psychology. Self-handicapping: The paradox that isn't". Plenum Press. 187-238

Jurjako, M. (2013) *Self-Deception and the Selectivity Problem*. Balkan Journal of Philosophy, 5 (2), 151-162

Hirt, E. & McCrea, S.M. (2001). The Role of Ability Judgments in Self-Handicapping. Personality and Social Psychology Bulletin, 27, 1378-1389

Lazar, A. (1997). *Self-Deception and the Desire to Believe*. Behavioral and Brain Sciences, 20 (1), 119-120.

Lynch, K. (2017). *An Agentive Non-Intentionalist Account of Self-Deception.* Canadian Journal of Philosophy, 47 (6), 779-798.

Marcus, E (2019). *Reconciling Practical Knowledge with Self-Deception*, Mind, 128 (512), 1205–1225

Martin, K. A., & Brawley, L. R. (2002). *Self-Handicapping in Physical Achievement Settings*: *The Contributions of Self-Esteem and Self-Efficacy*. Self and Identity, 1(4), 337–351

McCrea, S. M. (2008). *Self-Handicapping, Excuse Making, and Counterfactual Thinking: Consequences for Self-Esteem and Future Motivation*. Journal of Personality and Social Psychology, 95(2), 274–292

McCrea, S. M., & Flamm, A. (2012). *Dysfunctional Anticipatory Thoughts and the Self-Handicapping Strategy*. European Journal of Social Psychology, 42(1), 72–81.

Mele, A., & Moser, P. (1994). *Intentional Action*. Noûs, 28(1), 39-68.

Mele, A. (1997*). Real Self-Deception*. Behavioral and Brain Sciences, 20(1), 91–136.

Mele, A. (2001). *Self-Deception Unmasked*. Princeton University Press.

Mele, A. (2019). *Self-Deception and Selectivity*. Philos Stud.

Mercier H, Rolison JJ, Stragà M, Ferrante D, Walsh CR, Girotto V. (2017). *Questioning the Preparatory Function of Counterfactual Thinking*. Mem Cognit, 45(2):261-269.

Pears, D. (1984). *Motivated Irrationality*. Oxford University Press.

Quattrone, G. A., & Tversky, A. (1984). *Causal Versus Diagnostic Contingencies: On Self-Deception and the Voter's Illusion*. Journal of Personality and Social Psychology, 46(2), 237–248.

Porcher, J. (2012). *Against the Deflationary Account of Self-Deception*. Humana.Mente. 20, 67-84.

Rorty, A. (1994). *User-Friendly Self-Deception*. Philosophy, 69 (268):211 - 228.

Snyder, C.R., Higgins, R.L. (1988) *From Making to Being an Excuse*: *An Analysis of Deception in Verbal/Nonverbal Behavior*. Journal of Nonverbal Behavior 12, 237–252

Self, E. A. (1990). *Situational Influences on Self-Handicapping*. In R. L. Higgins (Ed.), “The Plenum series in social/clinical psychology. Self-handicapping: The paradox that isn't”. Plenum Press. 37-68

Smith, T. W., Snyder, C. R., & Handelsman, M. M. (1982). *On the Self-Serving Function of an Academic Wooden Leg*: *Test Anxiety as a Self-Handicapping Strategy*. Journal of Personality and Social Psychology, 42(2), 314–321.

Snyder, C. R., Smith, T. W., Augelli, R. W., & Ingram, R. E. (1985). On the self-serving function of social anxiety: Shyness as a self-handicapping strategy. Journal of Personality and Social Psychology, 48(4), 970–980.

Snyder, C.R. (1989). *Self-Handicapping Behavior and the Self-Defeating Personality Disorder*. In “Self-Defeating Behaviors”, Curtis. R (Ed.). Plenum Press

Van Leeuwen, N. (2013). *Self-Deception*. International Encyclopedia of Ethics, H. Lafollette (Ed.)