Self-Deception as a Moral Failure

Jordan MacKenzie
(Forthcoming at *Philosophical Quarterly*)

Abstract

In this paper, I defend the view that self-deception is a moral failure. Instead of saying that self-deception is bad because it undermines our moral character or leads to morally deleterious consequences, as has been argued by Butler, Kant, Smith, and others, I argue the distinctive badness of self-deception lies in the tragic relationship that it bears to our own values. On the one hand, self-deception is motivated by what we value. On the other hand, it prevents us from valuing those things properly. I argue that we owe it ourselves to take seriously our own values, by striving to properly value them. This gives us both prudential and moral reasons to avoid self-deception.

Keywords: self-deception, deception, values, valuing, valuer, self-knowledge, duties to self, epistemic vice

Word count: 9314

Introduction

Florence Foster Jenkins was a New York City socialite renowned for two things: her love of music and her complete lack of musicality. This lack of musicality was apparent to everyone, with the unfortunate exception of Florence. Indeed, Florence fancied herself to be an accomplished operatic soprano, and was known to 'treat' audiences to her tuneless warbling in private recitals throughout the 1920's, 30's, and 40's. Her 'career' culminated in a public concert at Carnegie Hall, where her performance caused uproarious laughter and pandemonium. Florence's lack of self-awareness was not the result of simple musical ignorance. She was, in fact, an accomplished pianist who had no difficulty judging *other peoples'* performances. Rather, the historical record paints Florence as a self-deceiver who maintained her self-deception by surrounding herself with a chorus of flatterers, and by dismissing the awful reviews of her recitals as mere professional jealousy (Bullock 2016).

There is clearly *something* undesirable about Florence's situation. At the very least, I know that I would strongly prefer not to be like Florence when it comes to any of my life projects. And we can

motivate this intuition from another direction, by considering the sorts of reactions we'd expect *Florence* to have if she were ever to discover her self-deception. When we uncover our own self-deceptions, after all, we typically react to these discoveries with a mixture of embarrassment, shame, guilt and self-contempt. Being self-deceived seems an intrinsically bad and undesirable state of affairs. But exactly what is so bad about self-deception? This is the question that this paper seeks to answer.

Historical accounts of self-deception's badness are a bit too heavy-handed for cases like Florence's. Bishop Joseph Butler, for instance, condemned self-deception as "a corruption of the whole moral character in its principle" (1729), while Adam Smith decried it as "the source of half the disorders of human life" (1759), and Immanuel Kant warned of it being the means through which our propensity for evil takes root within our moral character (1793; 1797). These diagnoses appear apt when applied to *some* self-deceivers—for instance, high-ranking Nazis who insisted that they had no idea what was happening in the concentration camps—but they seem overly moralistic when leveled at an out of tune songstress whose self-deception remained contained within a fairly narrow epistemic domain, and never resulted in any great moral harms.

More recently, there has been a resurgence of philosophical interest in self-deception, but with the exception of some neo-Kantian approaches (Darwall 1988; Baron 1988; Bagnoli 2012), the attention has largely shifted away from moral issues and towards the psychological and epistemic puzzles surrounding the phenomenon (see e.g. Fingarette 1969; Pears 1984; Mele 2001; McLaughlin and Rorty eds. 1988).[1] At the same time that these new debates have emerged, the classic moralistic viewpoint has come to seem old-fashioned, with some philosophers now arguing that self-deception is not as morally heinous as Kant and others envisaged it to be (Martin 1986; Kirsch 2005). Indeed, self-deception may well be personally beneficial (e.g. Rorty 1994; Kirsch 2005; Blumenthal-Barby and

[1] To illustrate this trend, note that the Stanford Encyclopedia article on self-deception includes only a short discussion of its moral dimensions (DeWeese-Boyd 2016).

Ubel 2018). While this optimism is not entirely misplaced, I still think it leaves out something important about Florence's situation. Her self-deception may have caused nobody any significant harm, but it still strikes me that her life has gone badly wrong in an important way.

On my view, Butler, Smith and Kant correctly sensed that self-deception involves a moral failing of some kind. But corruption of one's general moral character is not the fundamental issue with self-deception. What is bad about self-deception, I argue, is that it leads us to disrespect our own values. This is a moral failing because, as I will show, we are under an obligation to respect our own values, in much the same way that we are under an obligation to respect other peoples' values.

To show why self-deception's connection to valuing makes it a morally bad state that we have a self-regarding obligation to avoid, I first argue for the following claim in *Section 1*:

MOTIVATION: Self-deception is motivated by what we value.

That is, self-deception does not strike at random, but instead infiltrates the very epistemic domains that we, as valuers, ought to care about. My aim in arguing for **MOTIVATION** is not to provide a full account of self-deception's etiology, but instead to show why understanding self-deception's substance can help to explain why it is undesirable even in cases where it leads to no serious moral harms.

With this partial origin story in place, I move on in *Section 2* to argue for **RATIONAL PRESSURE**:

RATIONAL PRESSURE: If we value something, we are under rational pressure not to be self-deceived about it.

Self-deception, as I will show, undermines our ability to properly value the values that motivate it. Insofar as valuing something entails a commitment to properly valuing it, self-deception will be something that we are under rational pressure to avoid.

In *Section 3*, I go beyond this claim about rational failures, and argue for a claim about self-deception's moral badness:

MORAL FAILURE: Self-deception is a distinctly *moral* failure, because we owe it to ourselves to avoid undermining our own values.

To argue for **MORAL FAILURE**, I show that we, as valuers, have a *prima facie* self-regarding obligation to take seriously our values by (among other things) striving to properly value them. I wrap up this section by considering the limits of this moral obligation.

This leaves me with the following position on self-deception's moral badness. Self-deceivers ostensibly affirm a contradiction: their self-deception about some topic, X, affirms that they 1) think that X is worth valuing, and thus properly valuing and that 2) X is not worth valuing properly. But they haven't affirmed just any contradiction—they've affirmed a contradiction within an epistemic domain that they, as valuers, have committed to getting right. In failing to honor this commitment, they fall short of what they owe to themselves as valuers.

Section 1: Self-Deception and Valuing

My aim in this section is to argue for **MOTIVATION**, which says that self-deception is motivated by what we value. As will become clearer in the next two sections, I think that understanding what motivates self-deception can shed light on why we have reason to avoid it.

Before arguing for **MOTIVATION**, I'll say something about the epistemological literature to which it is responsive. Note that my aim is not to substantively engage with this literature, or to offer a complete account of the nature or origins of self-deception. In defending **MOTIVATION**, I hope to offer an account that is broadly parsimonious with the leading positions within the epistemological literature on self-deception. There are two broad types of answers that have been given to the question of how we become self-deceived. First, *intentionalist accounts* model self-deception after interpersonal deception, such that there is some sort of division within the self that allows us to intentionally deceive ourselves (Rorty 1972; Pears 1984; Davidson 1982; Bermúdez 2000). Second, *anti-intentionalist accounts*

understand self-deception as a process of desire-driven false belief-formation (Mele 2001; Johnston 1988; McLaughlin 1988; Barnes 1997; Nelkin 2002). The self-deceiver desires that some proposition, *p*, be true, and her desire leads her to "manipulate...a datum or data relevant, or at least seemingly relevant, to the truth value of *p*" such that she acquires the belief that *p* (Mele 2001, p. 51).

MOTIVATION, as I understand it, is consistent with both the anti-intentionalist picture and the intentionalist picture. If self-deception is intentional, we must presumably have some reason for deceiving ourselves. So in the intentionalist context, **MOTIVATION** is the claim that what we value supplies that reason. If self-deception is desire-driven, on the other hand, **MOTIVATION** is the claim that the type of desires that lead to self-deception are those desires that are tied up with our values. I will confess that my own sympathies lie with the anti-intentionalist, so I will be couching my argument in broadly anti-intentionalist language — with values playing the role that anti-intentionalists typically attribute to desires. But the basic point I am making is consistent with both views.

To explain why we gain something by identifying *valuing* rather than *desiring* as the motivator of self-deception, consider the following example. When I come home after a long day at work, I often find myself with a fairly strong desire to watch a saccharine reality television show like *The Great British Bakeoff*, and a much weaker desire to watch an acclaimed art film like *The Last Year at Marienbad*. My desire to watch *Bakeoff* has never led to any self-deception—if I don't watch the show one night, I don't find myself thinking that I'll *definitely* watch it the next night. My *Marienbad* desire, in contrast, is almost certainly the subject of self-deception. I sometimes, for instance, find myself musing that if even if I pass on *Marienbad* tonight, I will certainly watch it soon. Somehow, I still believe this, despite the fact that I have been putting off watching *Marienbad* for over half a decade.

Why am I self-deceived about my chances of watching *Marienbad*, but not *Bakeoff*? What distinguishes desires that motivate us towards self-deception from desires that don't, I believe, is the extent to which we identify with them. I *want* to be the sort of person who can appreciate art films—

and so I value having a desire to watch *Marienbad*. In contrast, while I also want to watch *Bakeoff*, I don't particularly value or identify with that desire—I don't take great pride in watching *Bakeoff* regularly, nor would I particularly care if I stopped desiring to watch it.

Self-deception, I want to suggest, bears a robust connection to what we value, and to the valuing process more generally. We can start to understand this connection by taking a closer look at what it means to value. By 'valuing', I have in mind a concept that is more selective than desiring; we may value many of the things we desire, but we may also desire things that we do not value, or that we positively disvalue. Valuing, further, is not reducible to any single attitude, belief, or behavior. Instead, it has been understood by Samuel Scheffler and others as a "complex syndrome" of doxastic attitudes, affective attitudes, actions, and commitments (2011; see also Svavarsdöttir 2014; Anderson 1993). On Scheffler's account, for instance, valuing something consists in judging it to be valuable or worthy, having affective and cognitive attitudes towards it that one regards as apt, and seeing it as generating reasons for action (2011, p. 32).

The "complex syndrome" that Scheffler describes can break apart in various ways. Most commonly, we find ourselves confronting gaps between the doxastic attitudes associated with valuing, and the action that those attitudes are meant to generate. Consider again the *Marienbad* example. When I say that I value art films, I don't simply mean that I think that it is a good thing that they exist in the world. Instead, my statement implies that I aspire to be a *valuer* of art films — which involves engaging with and appreciating them. What I lack, as a valuer, is action—I don't, as a matter of fact, actually watch many art films.

Similarly, we sometimes find ourselves acting in ways that reflect values that we *do not* endorse. Imagine that you take pride in your rigorously proletarian tastes, but nevertheless find yourself inclined to watch as many *Criterion Collection* films as you can get your hands on. You might feel all sorts of affective attitudes in relation to these films—excitement upon finding a DVD of *Marienbad* at a garage

sale, disappointment when the art house cinema in your town closes, and so forth. Nevertheless, you are embarrassed about these attitudes, finding them objectionably pompous. You live as though you value art films, but you don't identify as a valuer of them.

Do either of us value art films? There is a sense in which we both do—after all, we both tick off *some* of the boxes that Scheffler and others have associated with valuing. I have doxastic and affective attitudes, while you have affective attitudes and actions. Further, it would be reasonable for people to call both of us valuers. That someone could aptly accuse me of not living in accordance with my values, for instance, speaks to the fact that I value art films. Likewise, you could reasonably be charged with valuing art films more than you'd like to admit. But the senses in which we count as valuers are both far from ideal. I would be a better valuer of art films if I actually watched and enjoyed art films, and you would be a better valuer of art films if you endorsed your secret hobby.

For the purposes of my paper, I'm going to break valuing's 'complex syndrome' into two broad 'parts'. First, valuing can be identified with certain doxastic attitudes—believing X to be valuable, believing that it would be a good thing if one were a valuer of X, and believing that one ought to set ends with regard to the appreciation and promotion of X. Call this component of valuing '**endorsement**'. Endorsement typically also involves some affective attitudes; an endorser of X may *want* to set ends with regard to X, and may feel emotionally invested (to some extent at least) in being a valuer of X. Note that endorsement is the central sense of valuing that I will be working with in this paper. As such, whenever I refer to '**valuing**' *tout court* in this paper, I will be referring to valuing as **endorsement.**

The second part of valuing, which I will call '**embodiment**', involves treating values as reasons for actions, and setting ends in accordance with them. Embodiment, like endorsement, is typically accompanied by affective attitudes; I may enjoy pursuing the ends I set with regard to X, may feel frustrated when that pursuit goes awry, and so forth.

Often, we end up living in a way that embodies the values that we endorse. Call this alignment '**proper valuing**'. Sometimes, however, endorsement and embodiment come apart. We can fall short of proper valuing in at least two ways. First, we can embody values that we absolutely *don't* endorse, as illustrated in the case of the art film lover who nevertheless eschews all things pretentious. Second, we can incorporate our endorsed values into our lives in a way that does not appropriately reflect the value we take them to have. To illustrate: suppose that you, like Florence, really value being a good singer. This might motivate you to take vocal lessons, but it might also motivate you in another direction—you might find yourself so horrified by the prospect of being a bad singer, that you actively avoid situations in which you might be asked to sing. Active engagement and anxious avoidance both reveal our endorsed values. Nevertheless, they are not both ways of embodying our values. Indeed, anxious avoidance often ensures that we *never* embody our endorsed values.

With this picture of valuing, we can now establish **MOTIVATION**, which says that self-deception is motivated by our values—or more specifically, by the values we endorse. To start, note that self-deception *reveals* what we value. We can't even begin to understand Florence's self-deception, for instance, until we appreciate the value she places in being a good singer. Likewise, our endorsed values are revealed in cases of self-deception that do not involve beliefs about the self. Consider the sort of people who are self-deceived about the realities of global warming. Their self-deception speaks to their endorsed values—most typically, it speaks to the value they place in their conservative political identity. They endorse pro-business, anti-regulatory values, and they are resistant to evidence and beliefs that they perceive as threatening those values. Finally, we can even extend this valuing story to cases of what Alfred Mele (1999) has called 'twisted' self-deception (i.e., cases in which we have self-deceived beliefs about things being *worse* than they really are). The hypochondriac who self-deceptively thinks that every headache and freckle signals cancer is revealing something about what she values—namely, that she values being healthy.

We can also motivate **MOTIVATION** by considering how it stacks up against a closely related, but importantly distinct, way of understanding the connection between self-deception and valuing. Someone suspicious of **MOTIVATION** might argue self-deceivers are valuing *something else* which may be superficially similar to the value that they purport to care about, but ultimately distinct from it.[2] They might say instead that Florence doesn't *really* value being a good singer, but merely values being a *famous* singer. In the same spirit, we might tell a friend who is self-deceived about his unsatisfying relationship that he doesn't *really* value his romantic partner (much as he may say he does), but instead merely values the security that comes from having a long-term relationship.

But this kind of rationalized re-interpretation of the self-deceiver's behavior gets things wrong. To see why, remember what it means to embody a value. Embodied values are the values that we live by. Someone who *merely* embodies a value is thus living in a way that reflects a value she does not endorse. We see this discrepancy in the example of the person who is accused of being self-deceived about the fact that he is staying in his fraught relationship because he values security, and not because he values the person with whom he shares his relationship. When we accuse him of valuing security instead of his partner, we're not saying something about the values he endorses—we're saying something about the values he embodies.

So far, the objection gets something right about this case: the security-lover *does* value security. Nevertheless, the sense in which he values security is one of embodiment, rather than endorsement. That he is self-deceived about this embodiment, however, is evidence that he has a value that is both motivating his self-deception, and being undermined by it. Indeed, that the security-lover can't consciously reckon with the fact that he would rather stay in an unsatisfying relationship than subject himself to the vagaries of casual dating speaks to how much he *disvalues* being the sort of person who has that preference. He *wants* to be someone who only stays in a relationship if he loves the person

[2] I owe this objection to Philip Yaure.

with whom he shares it. This is the value he endorses, if not embodies. And so, he is self-deceived about staying with his partner—he thinks he's staying with her out of love, when really, he's staying with her out of habit. His self-deception thus keeps him in a situation in which his endorsed value will remain unsatisfied. So long as his fear of the unknown keeps him in his unhappy relationship, he will never find himself in a relationship that he'd want to stay in for reasons other than security.

To put this point another way, we must realize that self-deception has a protective function to play within our psychological lives, insofar as it shields us from painful realities. But these realities are only painful if we value their opposites. It was no great tragedy for me to realize that I can't sing, because I don't particularly care about singing. It is only a tragedy for Florence because she values being a good singer. If she ever abandoned this endorsed value, her self-deception would be unnecessary, and might (with sufficient evidence) disappear.

So self-deception reveals our values. But why think that these values are what motivate us towards self-deception, as **MOTIVATION** implies? To start answering this question, let's consider why self-deception tends to cluster around 'valuing domains'. We're self-deceived about our relationships and political affiliations, but not about our socks or whether we'd prefer peanut butter or Nutella.[3] I think that the fact that self-deception tends to concentrate within epistemic domains that implicate our values (e.g. our relationships, our children, our moral characters) can be partially explained by the type of thing that valuing *is*. To see why the process of valuing creates an especially fertile ground in which self-deception can grow, think back to the attitudes and dispositions that Scheffler associated with valuing. When we value something, X, we have an antecedent commitment to X's valuableness, an emotional vulnerability towards X that we see as apt, and a disposition to treat X-related considerations as reasons for action. These attitudes and dispositions, taken together, can give rise to

[3] To the extent that we can imagine being self-deceived about a Nutella preference, it will be because that preference bears some connection to our values. If I value a healthy diet, for instance, I might be self-deceived about how strongly I prefer Nutella to natural peanut butter.

self-deception. Antecedent beliefs about X's value can lead us to discount the testimony of people whom we perceive as not valuing X. Emotional vulnerabilities towards X can bias our evidence-gathering by making certain true beliefs around X too painful to countenance, and certain false beliefs too comforting to resist. And ongoing engagement with X can give us opportunities to selectively gather evidence about it, thus solidifying our self-deceived beliefs.

Finally, the very fact that valuing involves numerous attitudes, beliefs, and reasons for action can shed light on why it predisposes us to self-deception. Self-deception, I think, often gets its foothold in the gap between endorsement and embodiment. We might think that something is incredibly valuable, but nevertheless struggle to embody it in our lives. Or we might have emotional reactions in connection to the objects of our values that we find inappropriate or overblown. These sorts of mismatches are uncomfortable. Self-deception, then, can be understood as a way of bringing a feeling of stability to a valuing process that, due to its multifaceted nature, is prone to being unstable.

The preceding discussion should not be understood as providing a complete etiology of self-deception; such an account is well beyond the scope of this current project, and may not even be possible (Mele 2019). In particular, my account does not solve the 'Selectivity Problem' of self-deception. This problem is usually framed in terms of desires: if desires motivate us toward self-deception, why is it the case that not *all* desires (or even all unfulfilled desires) lead us to be self-deceived? (see e.g. Bermúdez 2000). There is an analogous question about why valuing something does not lead to self-deception in every case. What independent factor separates self-deceiving valuers from valuers who are not self-deceived? While this is a fascinating question, it is not one that I will attempt to resolve in this paper. Instead, I will now turn to the question of why self-deception's connection to valuing puts us under rational pressure to avoid it.

Section 2—Self-Deception and Improper Valuing

In this section, I will argue that when we value something, we are under rational pressure to avoid being self-deceived about it. I will argue for this claim as follows:

> *Premise 1* (from **MOTIVATION**): When we are self-deceived, we are always self-deceived about something we value.
>
> *Premise 2*: If we value something, we're under rational pressure to properly value it.
>
> *Premise 3*: If we're self-deceived about something, we're improper valuers of it.
>
> *Conclusion* (**RATIONAL PRESSURE**)*:* Therefore, if we value something, we are under rational pressure not to be self-deceived about it.

Section 1 defended *Premise* 1. In Section 2.1 I will argue for *Premise 2*. I then move on to argue for *Premise 3* by establishing that self-deception both causally contributes to devaluing (Section 2.2) and that it is constitutively incompatible with proper valuing (Section 2.3).

2.1—Valuing and Proper Valuing

To motivate Premise 2, recall first that proper valuing, on my account, involves an embodiment of one's endorsed values. But what embodiment involves is difficult to spell out in the abstract. This is partially because values themselves put constraints on what can count as proper valuing by making different ends, beliefs, and attitudes appropriate or inappropriate. For instance, properly valuing someone as a romantic interest isn't consistent with being their stalker. In addition, the particular ends that we set in regard to the objects of our values shape what is involved in properly valuing. Valuing music might lead you to learn an instrument, or to become a patron of your local orchestra—what it takes to properly value music will look different depending on which of these projects you pursue. We can set these nuances aside; all that I need for this argument is the idea that there are some ways of responding to our values that fall short of embodiment.

Regardless of what proper valuing involves in any particular case, I want to suggest that the fact that we value something puts us under rational pressure to properly value it. To see why, try to disentangle valuing from proper valuing. Suppose I told you that I cared deeply about the environment, and then continued to explain that I drove a gas-guzzling Hummer, proudly supported the illegal depletion of the rainforest, and would never donate to a conservation charity. What sense could you make out of my statements? You might think that I was somehow mistaken—perhaps, for instance, I'm mistaken in thinking that I value the environment. Or maybe you'd think I was just lying. Regardless, there would be something odd about what I have told you: specifically, my second statement seems to directly cut against my first.

My claim that valuing rationally requires a commitment to proper valuing should not be taken to imply that *any* deficiency in this commitment is tantamount to not valuing. One can be committed to proper valuing even if one sometimes falls short. Likewise, this commitment need not be categorically overriding. Still, even in cases in which we fall short of this commitment, we generally experience it as having normative force. An environmentalist might regret having to take a plane to visit a dying relative, for instance, even if they think that their decision was all-things-considered justified.

Valuing thus rationally requires a commitment to proper valuing in much the same way that setting an end rationally requires us to will the means to that end, or else abandon it. In the next two subsections, I will motivate Premise 3 by locating two ways in which self-deception leads us to disrespect, and thus fail to embody, the values motivating it: self-deception can *causally* contribute to us acting in ways that are anathema to proper valuing (2.2), and it can lead us to disrespect these values *constitutively*, such that the very state of being self-deceived is inconsistent with proper valuing (2.3).

2.2—Causal Devaluing

Let's first consider the causal link between self-deception and devaluing. As noted in *Section 1*, we set ends in relation to what we value. Self-deception typically makes us bad at figuring out what ends to set, and how to pursue our values. If you're self-deceived about how well you're doing academically, for instance, you won't see yourself as having a reason to study more.

Sometimes, self-deception might be a rather minor impediment to pursuing appropriate ends. At other times, it can lead us to 'pervert' the values that motivate it, such that we actually come to embody their opposites. To see what I mean by 'perversion' here, think back to Florence. Because she is self-deceived, Florence overestimates her ability to handle difficult repertoire. As a result, she turns serious arias into unintentional comedy numbers. Someone listening to Florence's rendition of the *Queen of the Night* aria for the first time would not think that she was simply a *bad* singer. Rather, they would likely assume that she was intentionally parodying the operatic form.[4]

This perversion is also present in more prosaic cases of self-deception. Suppose that you are so attached to your relationship that you can't seriously entertain the possibility that it may be in trouble. If you end up self-deceptively underplaying the seriousness of your relationship issues, your self-deception will be best explained by appeal to the value you place in that relationship: you wouldn't self-deceptively believe that your relationship was fine if you didn't care about continuing it. But just as your self-deception is motivated by the value you place in your relationship, so too does it undermine that value. If you can't see the cracks in your relationship, you won't be able to mend them.

Someone might object that self-deception can actually at times *promote* our ends, and the values they reflect (Szabados 1974, Kirsch 2005). It can, for instance, sometimes gives us the "fake it 'til we make it" confidence that we need to accomplish our ends. A fledgling comedian might need to be a

[4] A recording of the aria is available here: https://youtu.be/V6ubiUIxbWE

little self-deceived about her chances of bombing onstage in order to have the confidence to actually do well at open mic night. Her self-deception is not undercutting her values, but is instead actively promoting the ends that she has set with regard to them.

There are two things to say about this case. First, if one thinks that self-deceived beliefs are necessarily false, or at least epistemically unwarranted, then this case may not be a case of self-deception, provided the performance is in fact successful. The comedian arguably has a true and epistemically warranted belief—namely that her set will not bomb—where part of the evidence for the belief is the presence of the belief: the fact that she is confident because she thinks she won't bomb makes her less likely to bomb (McLaughlin 1988, p. 46). The comedian may not have relied on this particular piece of evidence to form her belief—but the evidence that she does access presumably can't have ruled out the truth of her belief. If it had, then *no* amount of confidence would have been sufficient for her to have a good set. That many low-level 'self-deceptions' don't actually involve false beliefs might be sufficient to make us think that they're some other species of motivated reasoning.

That said, there are certainly some fledgling comedians who have a completely *unwarranted* belief in their chances of succeeding at their first open mic night, such that no amount of confidence could possibly up their chances of not bombing. Here's what my account can say about them. Self-deceived comedians value being good comedians. Their willingness to perform at open mic nights, along with their self-deception towards their performances are both evidence of this fact. Because they value being good comedians, they're invested in properly valuing. As such, we can trust that they don't want to be delivering unfunny performances. And yet that's exactly what their self-deception leads them to do: to give a terrible, unfunny performance that is at odds with their endorsed value. In this sense, their self-deception helps make them into what they want to be least of all, namely, an unfunny comedian.

To be sure, giving some terrible performances may often be a necessary means to having a great set. But this doesn't erase the fact that those awkward initial performances, fueled by self-deception, were crimes against comedy. Even once a comedian becomes successful, they will likely still be embarrassed about those awful early shows. And they will further be embarrassed by the fact that they were self-deceived about how bad they really were at the beginning, insofar as their self-deception cut against their commitment to producing good comedy. In other words, even if they grant that it was all things considered a good thing that they were self-deceived at the beginning of their career, they might still reasonably wish that they could have found some way to be both self-aware and willing to go on stage. Thus, self-deception frustrates proper valuing even when it is fairly trivial, and when it is ultimately pragmatically beneficial.

Occasionally, there may be cases where sheer luck prevents self-deception from having an impact on the objects of our values, or where self-deception's causal devaluing is ostensibly 'cancelled out' by sheer luck or a countervailing prudential commitment. I might, for instance, be self-deceived about my chances of passing an exam, only to have the exam cancelled at the last minute. Or, the effects of my self-deceived belief that I am good at math might be tempered by my standing policy of always double-checking my sums (no matter how confident I am about them). But these cases are not counterexamples to the view; they are simply cases where the undermining effects of self-deception have been blocked by extrinsic factors. My self-deception in these cases is still value-undermining, even if sheer luck or prudential policy prevents its deleterious effects from becoming manifest.[5]

[5] Likewise, we can say that drunk driving is safety-undermining even in cases when it doesn't cause an accident.

2.3—Constitutive Devaluing

In addition to causally devaluing the values that motivate it, self-deception is *constitutively* at odds with proper valuing, such that the very state of being self-deceived makes impossible some components of proper valuing.

First, self-deception frustrates our ability to know the objects of our values, and the relationship that we presently bear to them. This violates what I take to be a constitutive requirement of proper valuing—if we value something, we should know, and should want to know, what it is that we're valuing and how we're doing at valuing it. Valuing someone as a friend or romantic partner, for instance, involves having some sort of ongoing interest in knowing more about them, as well as a commitment to knowing where we stand in relation to them (MacKenzie 2018). Our curiosity need not be unbounded (even our nearest and dearest sometimes bore us to tears), but its complete absence is generally a sign that our valuing has run its course. And so too does valuing projects, ideals, and objects of inquiry involve a commitment to knowing. If a purported valuer of philosophy were to tell you 'I value philosophy—but I don't know, or care to know, what it is', you'd think that she was pulling your leg.

That proper valuing involves a commitment to knowing does not imply that we need to know, or even want to know, *everything* about the objects of our values in order to be proper valuers. Further, valuing something often involves being antecedently predisposed to think well of it in a way that can be in tension with having maximally accurate beliefs. If I value you as a friend, I may be more likely to give you the benefit of the doubt when you do something hurtful. Instead, when I say that valuing involves a commitment to knowing, I mean that there are limits to the amount of ignorance and insensitivity we can have towards objects of value, and the ends that we set in regard to them. It's one thing to find certain facets of your life project uninteresting—and it's quite another to find the entirety

of that project dull. Likewise, it's one thing to see your child through rose-colored glasses, and it's another to not have the faintest sense of who they really are.

Self-deception is at odds with the commitment to knowing that is part of proper valuing. Thus, insofar as it keeps us from seeing clearly the objects of our values, and our relation to them, it is constitutively incompatible with proper valuing.

Self-deception is constitutively incompatible with proper valuing in at least one other way. Valuing, as previously noted, involves emotional vulnerability. It can be a great source of joy, but it can also be a great source of sorrow. Properly valuing something typically requires that we view both sorts of emotions as potentially apt reactions to our values. To make this aspect of valuing more vivid, think about the intense sorts of emotional reactions that one may experience in the process of trying to launch a music career. A fledgling musician might feel joy at having landed a gig, but she'll also feel sorrow if nobody comes to hear her play, and embarrassment if she performs poorly. These sorts of negative emotional reactions aren't regrettable byproducts of valuing. Rather, they're part of what is involved in being emotionally vulnerable to one's values.

Self-deception, in contrast, gives us a 'get out of jail free' card in relation to valuing's most painful emotions. Florence's self-deception protects her from feeling the disappointment and embarrassment that her cacophonic recitals would otherwise engender. But in avoiding this range of negative emotions, she cuts herself off from an important aspect of valuing. She loses out on the full valuing experience that those of us who aren't self-deceived get to partake in.

Section 3—Moralized Self-Deception

Let's recap. Self-deception is incompatible with proper valuing. And insofar as valuing rationally requires a commitment to proper valuing, it will be something that we are under rational pressure to try to avoid. Nevertheless, **RATIONAL PRESSURE** need not have much moral 'oomph' by itself. We

are, for instance, under rational pressure to not hold contradictory beliefs. But it doesn't follow from this that we have a moral obligation to spend lots of time trying to discern whether beliefs from disparate epistemic domains are fully compatible with each other. Further, when we discover such incompatibilities, we rarely have the sorts of moralized reactions to them that discoveries of self-deception typically engender. I don't feel that I've let myself (or anyone else) down when I realize that I both believe that Vatican City is a city in Italy, and that Vatican City is the smallest country in the world.

In this section, I aim to argue for **MORAL FAILURE**, which says that self-deception is a distinctly *moral* failure because we owe it to ourselves to avoid undermining the objects of our values (Section 3.1). I'll also consider the limits of this obligation (Section 3.2).

Note that I will be presupposing the existence of self-regarding obligations. Readers who are skeptical about self-regarding obligations are invited to translate my claims about them into claims about the moral reasons that we have to treat ourselves in certain ways.

3.1—Why We Have Moral Reason to Avoid Self-Deception

The reason that self-deception often strikes us as a moral shortcoming is that it undermines something that we have moral reason to take seriously: the objects of our values. In cases where these objects are morally relevant, we have a straightforward moral reason to take them seriously. Consider Carla Bagnoli's example of a mother who is self-deceived about her daughter's anorexia (2012, 99-100). The mother's self-deception leads her to violate the duties of care that she has towards her daughter. Her self-deception speaks to the fact that she values her daughter's wellbeing, and recognizes herself to have duties to promote it, but it also stands in the way of her helping her daughter.

Since we have moral reason to embody many of our endorsed values, self-deception is often something that we have a straightforward moral reason to try to avoid. But while this can explain the

moralized reactions that we would likely have to the mother who is self-deceived about her daughter's anorexia, it does little to explain the moralized reactions that we are likely to have in cases like Florence's. The value that Florence is disrespecting, after all, is not a moral value: if bad singing were a moral wrong, then karaoke bars would be dens of iniquity. And yet, I think it would be reasonable to think that Florence *ought* to know better and that she *shouldn't* go so easy on herself. Lest there is any doubt that these oughts and shoulds are moralized, consider the reactions we'd expect Florence to have towards her own self-deception if she ever uncovered it: we'd think it would be reasonable for her to experience some measure of shame, guilt and self-reproach.

How can self-deception instill non-moral values with moral content? The answer lies in the connection that valuers bear to their endorsed values. We are, fundamentally, *valuers*—we are not simply automata propelled around by whichever first-order desires happens to be strongest, but are instead agents who make judgments about what and how to value (Bratman 2000). What we value is a product of those judgments. A failure to take seriously our endorsed values is thus a failure to take seriously those judgments, and by extension, the person who made them.

To clarify this last point, consider the reasons that we typically have to take seriously *other peoples'* value judgments. First, we may have a reason to take their value judgments seriously because they are good judges. When my foodie friend recommends a restaurant, or my effective altruist friend recommends a charity, I take those recommendations seriously. Second, we may have a reason to take peoples' value judgments seriously simply because they're *theirs*—the very fact that you've decided to structure a significant portion of your life around some project (be it tap dancing, political activism, parenthood, and so forth) gives me a reason to take that project seriously. I don't have to approve of it or actively engage with it—but I shouldn't dismiss it out of hand, ridicule it, or condemn it (unless I have a good countervailing reason to do so).

I think the same reasons to take values seriously are present in the first-personal case as well. First, we typically have an expertise about ourselves that others lack (even if self-deception prevents perfect transparency). We thus have an authority-related reason to take seriously the judgments we've made about what it would be good for us to value. This reason is defeasible, but it provides a presumptive case in favor of taking seriously the values we endorse. Second, our status as autonomous, rational agents gives us a reason to take seriously our endorsed values. To constantly second guess our judgments about what values to endorse, or to not attempt to embody those values, is tantamount to not respecting ourselves as agents capable of making decisions about how to live.

We can see the requirement to take seriously the objects of our values by embodying them when we consider the extent to which people 'moralize' their life projects. An artist might feel that she owes it to her art to complete it, while philosophers (in my experience) sometimes talk about 'doing right' by their arguments. And someone who has embraced an 'alternative' personal aesthetic may think that she's sacrificed something morally important when she takes out her piercings to placate her boss. I do not claim that we *actually* have moral obligations to our paintings, philosophical projects, or piercings. But we do have moral obligations to ourselves as valuers—one of which, I believe, is to take the objects of our values seriously by striving to properly value them.

This explains why we have a moral reason to avoid self-deception even in 'non-moral' cases like Florence's. Self-deceivers ostensibly affirm a contradiction. Through their actions and attitudes, self-deceivers simultaneously affirm: (1) that X is worth valuing, and thus properly valuing and (2) that X isn't worth properly valuing. But they haven't affirmed any old contradiction—they've affirmed a contradiction within the very epistemic domain that they, as valuers, have antecedently committed to taking seriously. In failing to honor this commitment, they fail to take themselves seriously as valuers.

3.2 The Limits of Our Obligation to Avoid Self-Deception

I have argued that that self-deception is a distinctly *moral* failure because we owe it to ourselves to avoid undermining our own values. This does not mean that our obligation to avoid undermining our own values is unconditional: the moral reasons we have to properly value the objects of our values and thus to seek to avoid self-deception can certainly be outweighed by weightier countervailing moral considerations. In this section, I'll discuss two such considerations.

3.2.1—Immoral Values

First, consider cases in which the value driving our self-deception is immoral. Imagine that a self-deceived pickup artist believes himself to be a Lothario, despite actually being wildly unsuccessful with women. This case fits the picture of self-deception that I have provided: the value that this man places in being a pickup artist has led him to become self-deceived about his skill at seducing women. But it leaves us with two questions: is the self-deceived pickup artist disrespecting the 'value' of being a pickup artist by being self-deceived about the extent to which he has successfully embodied it? And is he thereby disrespecting himself by failing to properly value his values? The answer to these questions depends on whether one thinks that we can ever be obligated to properly value something that, morally speaking, ought not to be valued.

Let's suppose that we can be under such pressure—the very fact that one values being a pickup artist, in other words, would put one under rational (and on my account, moral) pressure to properly value this objectionable goal. So long as the self-deceived pickup artist retains a commitment to valuing pickup artistry, his self-deception will constitute a failure to properly value something that he has committed to properly valuing.

Of course, whatever moral badness results from this failure, it will be outweighed by competing moral considerations. Even if there's something bad about his self-deception, it may still

be consequentially good that his self-deception frustrates his efforts to achieve his objectionable goals. And we could also think that, while he may be obligated to take seriously his present values by properly valuing them, he's also obligated to make sure that his values are worth valuing.

Now let's suppose that we can't be under rational pressure to properly value something that, morally speaking, ought not to be valued. We could put this position in terms of properly valuing: perhaps there simply is no such thing as 'properly valuing' the 'value' of being a pickup artist because pickup artistry fundamentally rests on mistaken beliefs about how other people should be treated. If so, then our pickup artist is under rational pressure to abandon the value motivating his self-deception: since valuing rationally requires a commitment to proper valuing, and since that commitment is unfulfillable in this case, he ought to abandon the 'value'. Since his self-deception prevents him from responding to this pressure, he is falling short of what he owes to himself as a valuer. Under both interpretations, our self-deceived pickup artist is falling short—and it makes sense that he would react to a discovery of his self-deception with feelings of shame, guilt and self-reproach.

Both analyses of the case bring out another, more fundamental way in which self-deception frustrates us as valuers. Respecting ourselves as valuers requires that we take the objects of our values, whatever they are, seriously by properly valuing them. But it also requires that we take seriously the process of choosing what values to endorse by subjecting our value judgments to scrutiny. Since these values reflect on us as valuers, we should want them to reflect well on us.

Of course, we can't constantly scrutinize our endorsed values. And nor should we: if we did, we'd never get around to actually trying to embody them. This is why failures of proper valuing are important: they provide us with specific opportunities to reflect on what, and on how, we value. The person who drops out of college, or who can no longer stand to be around their partner is given a chance to ask: is this really worth valuing? And if so, is it worth valuing like *this*? Whether they decide

to double down and properly value, or move on to greener pastures, they are still afforded an opportunity to take themselves seriously as valuers.

Self-deceivers, in contrast, cut themselves off from the specific opportunities to reflect on these questions that valuing failures provide. Because he never realizes how bad he is with women, the wannabe-Lothario may not question the value of pickup artistry. And because she never realizes how bad a singer she really is, Florence may never question whether there is some other way to value singing, other than performing at Carnegie Hall. Self-deception thus frustrates our ability to embody the values we endorse, as well as our ability to question whether our endorsed values are worth embodying. So long as we owe it to ourselves as valuers not to fall short of proper valuing, our self-deception will constitute a failure to fulfill a self-regarding obligation.

3.2.2—Self-Deception Beyond Reproach

Consider now a case of self-deception that seems morally beyond reproach. It's one thing to say that Florence has wronged herself by being self-deceived, but it is quite another to argue that a late-stage cancer patient who self-deceptively believes that her prognosis isn't terminal has fallen short of what she owes herself. This judgment is reflected in the sorts of third-personal reactions we often have to such self-deceptions: such self-deception engenders empathy, not contempt, and might be something that we have reason to promote (or at least not frustrate) amongst terminally ill patients. (Blumenthal-Barby and Ubel 2018). Thus, there seems to be an important class of self-deceptions that we are not obligated to avoid.

I agree that the threat that the truth poses to our wellbeing is sometimes significant enough to warrant self-deception, and thus to justify an exception to the duty I have described. But this conclusion is unsurprising: the obligation we have to take seriously the objects of our values can at times be outweighed by competing moral considerations.

Even so, there is still *something* that the self-deceived patient loses out on. We might recognize that it is all-things-considered in his interests to be self-deceived, while still thinking that there would be *something* better about him facing his death with clear-eyed self-awareness. What the self-deceived patient loses out on, I think, is his last opportunity to properly value his own life. This value can be seen to be motivated by his self-deception—if he didn't care so much about living, he wouldn't be self-deceived about his own death. But it is also undermined by his self-deception; by being self-deceived, he will likely lose out on some of his last opportunities to engage fully with the projects and people who gave his life meaning. He will also lose out on something more fundamental—the ability to engage fully with his life as it really is. His self-deception, while on balance morally defensible, still comes at a cost. And it's a morally relevant cost, given what he values.

We can give a similar analysis for cases in which there appears to be something positively virtuous about being self-deceived. Consider Szabados's discussion of a mother who self-deceptively clings to the belief that her son is still alive, despite mounting evidence to the contrary (1974, p. 28-29). There is something morally admirable in this response, insofar as it speaks to how much the mother loves her son. This result is unsurprising on my account as self-deception speaks directly to what we value, and some of what we value is typically morally laudatory. But even in cases where there is something admirable about self-deception, it is still a failure of valuing. The self-deceived mother, for instance, cuts herself off from the sorts of painful, but apt, emotional responses that proper valuing would lead her to experience: she cannot grieve her son because she cannot see that she has reason to grieve. And in this way, she falls short of proper valuing. It is a failure that we should not condemn, and that we may even admire. But it is a failure nonetheless.

Conclusion

We can now explain why self-deception is morally bad even when it doesn't lead to any great moral harms. Its badness is locatable in the connection that it bears to the objects of our values: rather than striking at random, self-deception infects the very epistemic domains that we have antecedently committed to getting right. This means that we need not look outside the self to locate the moral badness of self-deception; instead, we need only to notice that self-deception involves a failure to properly relate to one's values, and thus to oneself as a valuer.

Acknowledgments

This paper has benefitted from a lot of feedback over the years since its creation. It would not be in the state that it is presently in without help from: Ben Bagley, Samuel Curtis, Daniel Fogel, Jonathan Gingerich, Thomas E. Hill, Rachel Keith, Adam Lerner, S. Matthew Liao, David Merli, Ram Neta, Claudia Passos-Ferreira, Chelsea Rosenthal, Susan Wolf, Philip Yaure, and two anonymous reviewers at this journal. Thanks also to audiences at Ashoka University, the Obligations to Oneself Workshop, the American Philosophical Association (Central Division), Uppsala University, McGill University, Virginia Tech, Fordham University, the CUNY Moral Psychology Reading Group, and the UNC Dissertation Research Seminar. Finally, thanks to Daniel Hoek, who never seemed to get sick of talking about this paper with me.

Works Cited

Anderson, E. (1993). *Value in Ethics and Economics*. Cambridge, MA: Harvard University Press.

Bagnoli, C. (2012). Self-Deception and Agential Authority: A Constitutivist Account. *Humana.Mente Journal of Philosophical Studies* 20: 99-116.

Barnes, A. (1997). *Seeing Through Self-Deception*, Cambridge, UK: Cambridge University Press.

Baron, M. (1988). What Is Wrong with Self-Deception? In B.P. McLaughlin and A.O. Rorty (Eds.), *Perspectives on Self-Deception*. Berkeley, CA: University of California Press.

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

Blumenthal-Barby, J. and P. Ubel. (2018). In Defense of "Denial": Difficulty Knowing Whether Beliefs are Unrealistic and Whether Unrealistic Beliefs are Bad. *American Journal of Bioethics*, 18(9): 4-15.

Bratman, M.E. (2000). Valuing and the Will. *Philosophical Perspectives, 14, Action and Freedom*: 249-265.

Bullock, D.W. (2016). *Florence! Foster!! Jenkins!!!: The Life of the World's Worst Opera Singer*. New York, NY: The Overlook Press.

Butler, J. (1729). Sermon X: Upon Self-Deceit. Virginia: Lincoln-Rembrandt Publishing, 1993.

Darwall, S. (1988). Self-Deception, Autonomy, and Moral Constitution. In B.P. McLaughlin and A.O. Rorty (Eds.), *Perspectives on Self-Deception*. Berkeley, CA: University of California Press.

Davidson, D. (1982). Deception and Division. In J. Elster (ed.), *The Multiple Self*. Cambridge, UK: Cambridge University Press, pp. 79-92.

DeWeese-Boyd, Ian. (2016). *Self-Deception*. Stanford Encyclopedia of Philosophy.

Fingarette, H. (1969). *Self-Deception*. Berkeley: University of California Press.

Kirsch, J. (2005). What's so Great about Reality? *Canadian Journal of Philosophy* 35(3):407-427.

Kant, I. (1797). *Metaphysics of Morals*. In I. Kant, *Practical Philosophy* (pp. 37-108). Tr. and Ed. M.J. Gregor. *The Cambridge Edition of the Works of Immanuel Kant*. Cambridge, UK: Cambridge University Press, 1996.

------- (1793). *Religion Within the Boundaries of Mere Reason*. Tr. And Eds. A. Wood & G. Di Giovanni. Cambridge: Cambridge University Press, 2012.

Johnston, M. (1988). Self-Deception and the Nature of the Mind. In B. P. McLaughin and A. O. Rorty (Eds.), *Perspectives on Self Deception*, Berkeley: University of California Press, pp. 179-88.

MacKenzie, J. (2018). Knowing Yourself and Being Worth Knowing. *Journal of the American Philosophical Association*, 4(2): 243-261.

Martin, M. (1986). *Self-Deception and Morality*, Lawrence: University of Kansas Press.

McLaughlin, B.P. (1988). Exploring the Possibility of Self-Deception. In B. P. McLaughin and A. O. Rorty (Eds.), *Perspectives on Self Deception* (pp. 29-62). Berkeley, CA: University of California Press.

Mele, A. (1999). Twisted Self-Deception. *Philosophical Psychology,* 12: 117-137.

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

------- (2019). Self-Deception and Selectivity. *Philosophical Studies.*

Nelkin, D.K. (2002). Self-Deception, Motivation, and the Desire to Believe. *Pacific Philosophical Quarterly* 83: 384-406.

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

Rorty, A.O. (1972). Belief and Self-Deception. *Inquiry*, 15: 387-410.

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

Scheffler, S. (2011). Valuing. In R. J. Wallace, R. Kumar, and S. Freeman (Eds.), *Reasons and Recognition: Essays on the Philosophy of T. M. Scanlon* (pp. 24-39). Oxford: Oxford University Press.

Smith, A. (1976). *Theory of Moral Sentiments*. D.D. Raphael and A.L. Macfie (Eds.). Oxford: Oxford University Press.

Svavarsdöttir, S. (2014). Having Value and Being Worth Valuing. *The Journal of Philosophy*, 111(2); 84-109.

Szabados, B. (1974). The Morality of Self-Deception. *Dialogue: Canadian Philosophical Review*, 13 (1): 25-34.