Aporia vol. 28 no. 1–2018

The Problems, Paradoxes, and Epistemic Implications of Self-Deception

Josephine Lovejoy

In his essay entitled "Self-Deception and Rationality," Robert Audi suggests three criteria that categorize a mental state as self-deception. He explains that a person S is in a state of self-deception with respect to a proposition p if she:

> 1. Unconsciously knows that ¬p.
> 2. Sincerely avows that p.
> 3. Feels the desire to explain:
>
> a. Why ¬p is unconscious, and why she is inclined to disavow ¬p.
>
> b. Why she is inclined to avow p even in the face of evidence against p. (Audi 73)

To illustrate these criteria, I will introduce an example that I will refer to for the remainder of the discussion. Violet is a passionate vegetarian and animal rights advocate and believes that it is immoral for humans to purchase and wear animal fur.[1] While shopping at a thrift store on a cold winter day, Violet spots a beautiful fox fur coat. She realizes that the coat is made of real animal fur but decides to purchase it anyway. Violet deceives herself into believing that it is *not* immoral to purchase and wear animal

[1]I will refer to this belief as "¬pV."

Josephine is a senior at at Vassar College and a double major in mathematics and philosophy. After graduation, she plans to attend law school with a focus on either legal academia or intellectual property law. Her philosophical interests include logic, philosophy of mind, and philosophy of law.

fur.[2] On Audi's model, we may characterize this scenario as an example of self-deception if we can ascribe to Violet the following characteristics. First, she unconsciously knows that ¬pV.[3] Then, she sincerely avows that pV. Further, Violet's self-image as a vegetarian is confronted with a piece of contradictory evidence upon purchasing the coat. Violet then experiences an uncomfortable tension between her act of purchasing the coat and her belief that it is immoral to purchase or wear animal fur. Her desire to eliminate this tension explains why she might feel the need to avow pV. And, because her desire to eliminate this tension is so strong, Violet is also likely to insist that she believes pV, even in the face of evidence as to the contrary.

The model of self-deception outlined above and illustrated through Violet seems to pose an apparent problem. Consider the interpersonal case of deception. When we say that person A deceives person B with respect to a proposition p, we say that person A somehow convinces person B to believe that p, when all the while person A knows that p is false. What is peculiar in *self*-deception as Audi points out is that in the same person we find both the deceiver and the deceived (171). Thus, our person S should both unconsciously believe that ¬p and consciously believe that p. But to believe both a proposition p, and the opposite of that proposition, ¬p, is a logical contradiction—one cannot believe both (Audi 172)! That is, Violet cannot reasonably believe that her purchasing the fur coat was both moral and immoral. If this curious paradox cannot be solved, we are left to wonder what we really mean when we say that we have "deceived ourselves." Is the concept of self-deception an empty one?

Audi offers a resolution to this paradox. He explains that "while it is as *if* S believed what [she] knows is not true, self-deception stops just short of this" (Audi 175). That is, just because S is genuine in her avowal that she believes p, sincere avowal does not imply belief. Returning to Violet, she may sincerely avow that her actions were not immoral, but her sincerity here does not prove that she wholeheartedly believes this avowal. Thus, the paradox is solved: S does not *believe* both p and ¬p. However, her sincere avowal of p and her unconscious belief that ¬p justifies us in saying that S "deceived herself" (Audi 175).

I am unconvinced by Audi's resolution. While it does eliminate the paradox, it also seems to reduce the force of what we take the word

[2]I will refer to this belief as "pV."

[3]I do not wish to claim that wearing fur is objectively immoral, but rather that for Violet, it is contrary to her vegetarianism and therefore immoral f*or her.*

"deception" to be. That is, if rather than *believing* p S only *sincerely avows* p, is S really deceiving herself? Let's look to what Audi says. He writes that with respect to the proposition p, "S does not literally believe it, though it is natural to say that [she] 'consciously believes' it, and perhaps [she] may be said to 'half believe' it" (Audi 173). But neither sincere avowal, half belief, nor conscious belief seem to imply that S is truly deceived. Recall the previously explained interpersonal case of deception. In that case, we would not claim that person A has succeeded in deceiving person B if person B only half-believes, or even sincerely avows p, because there is still part of B that knows that p is false. Analogously, if S doesn't truly believe that p, we cannot claim that S has succeeded in deceiving herself. In the example of Violet, suppose that she only "half-believes" or even "is trying very hard to believe" that purchasing the coat was not immoral. If she can only offer these weakened claims of belief, it seems we cannot honestly claim that her attempted deception was successful.

The present claim is that Audi's resolution to the paradox is unsatisfying at best. But perhaps the reader is convinced by his resolution, and if so, the question I ask is this: If her belief in ¬p is unconscious, from where does S get the conscious idea that this belief is undesirable, i.e., in conflict with her recent behavior? In other words, if S is unaware of her belief in –p, how is she able to know that it is the culprit of her felt tension? I will soon elaborate on these questions, but first I will offer a passage of Audi's discussion that may help to shed light on these concerns.

Audi explains that unconscious beliefs are "simply not accessible to the conscious mind without outside help, or at least careful self-scrutiny" (174). Now it is unclear as to how this would actually work, but first let's suppose that through self-scrutiny I *am* able to access my unconscious belief that ¬p; my belief in ¬p is then brought into my consciousness. But if I am now holding in my conscious mind my belief in ¬p, how am I able to sincerely avow my belief in p? It seems that for my avowal of p to be sincere, my belief in ¬p must remain unconscious (Audi 173). Returning to Violet, if she consciously believes that purchasing the coat is immoral, and then consciously avows that purchasing the coat is not immoral, it appears that her avowal cannot be genuine. To avow pV while all the while knowing that she believes ¬pV seems to be a fruitless endeavor.

Now let us suppose that through self-scrutiny S is *not* able to discover her unconscious beliefs. Audi still claims that S is able to form a conscious avowal that p. But in order for S to avow p, she must know that her belief in ¬p is in tension with her previous actions that threatened her self-image—this tension is, after all, the reason she feels the need to affirm p in the first place. Further, in order to know that her belief in ¬p is in tension with these actions, it seems as though S must *consciously* know that she believes

¬p: she cannot find tension between her actions and a belief that remains unknown to her (Gergen 232)! Yet it is unclear how S can consciously know that she believes ¬p if this knowing remains unconscious.

To summarize, it seems that in order for S to form the intention to avow p, on Audi's model she either a) has privileged access to her unconscious mind, i.e., is able to "look into it" on demand in order to inspect for beliefs that are disharmonious with her prior actions or b) her unconscious mind is able to recognize the wish of the conscious mind to avoid conflict in self-image, and alert the conscious mind that it needs to assert p,[4] all while remaining anonymous to the conscious mind. Option (a) requires us to accept questionable Freudian conceptions of the unconscious,[5] and option (b) leads us to the issue of *subception*, which is one of the key reasons that psychologist Kenneth Gergen, in his essay "The Ethnopsychology of Self-Deception," rejects self-deception as a legitimate mental state.

Gergen argues that if S is unaware of her unconscious belief that ¬p, but has the conscious belief that p, "[one] must posit a subconceiving agency operating below the level of conscious awareness, yet serving the interests of the conscious mind" (232). In other words, one is "logically pressed into developing yet another form of consciousness, one that perceives or registers the undesirable impulses of the unconscious, sets defenses in motion, but does not report its activities to conscious awareness" (Gergen 232). On Gergen's account, Audi would have to concede to the existence of an intermediary consciousness that:

> 1. Can view the contents of the unconscious mind.
> 2. Notices when a proposition in the unconscious mind is contrary S's recent behavior.
> 3. Feeds a corrective action—an avowal that p—to the conscious mind while still remaining anonymous to the conscious mind. (Gergen 233)

On Gergen's account, there is no evidence that such a subconveiving device exists (232). Further, the complexity of the functions we must ascribe to this device make its existence, and therefore self-deception as a mental state, all the more dubious (Gergen 232–33). Lastly, Gergen explains that psychological research strongly suggests that we do not learn about our mental states through some process of introspection, giving us little reason to believe that we can affirm the existence of self-deception through

[4]And when S does assert p, her self-image is adjusted accordingly to reflect this new belief.
[5]The Freudian characterization is Kenneth Gergen's terminology, see Gergen 232.

personal observation and/or experience (234–35). These conclusions lead Gergen to suggest that "Self-deception is a constituent of the culture's ethnopsychology, or system of folk beliefs about the nature of human functioning at the psychological level (236). That is, it seems that we have taken our understanding of interpersonal deception and turned it inward to form our notion of self-deception.

Aliefs: A Possible Mechanism for Self-Deception?

Allow me to recap what I have found thus far. I outlined Audi's theory of self-deception and found difficulties in its requirement that S avow p consciously yet believe ¬p unconsciously. Such difficulties led Gergen to conclude that self-deception exists only as we use it in common discourse, but that it is unachievable as a mental state. Are we to accept this claim?

Gergen finds that the subception objection effectively destroys the legitimacy of self-deception as a mental state, but perhaps he is too quick in his judgment. What I have in mind here is the concept of "aliefs" as discussed in Michael Brownstein and Alex Madva's essay "The Normativity of Automaticity." By now we are familiar with beliefs; aliefs, then, are mental "states that dispose agents to respond automatically to apparent stimuli with certain fixed affective responses and behavioral inclinations" and "are causally responsible for the brunt of moment-to-moment behavior" (Brownstein and Madva 412). Take for example the male CEO who explicitly endorses the belief that male and female employees should be treated equally, but who implicitly endorses the idea, or "alieves," that they should not be treated equally.[6] This alief causes the CEO to think and behave in a sexist manner, e.g., become automatically frustrated when a female employee critiques his leadership skills.

Based on this idea, is it possible that aliefs, rather than unconscious beliefs, can do the work of self-deception? As Brownstein and Madva explain, a core feature of an alief in "good standing" is that "its motor and affective components work in concert to reduce 'felt tensions,' or experiences of 'disequilibrium between an agent and her environment" (412). To elucidate this idea, Brownstein and Madva provide the example of a museumgoer who, as she contemplates a large painting, feels the need step back and reorient her body to better view the painting (417). What she has done is eliminate a felt tension between the orientation of her body

[6]Note that aliefs need not always have specific content or reflect certain desires—they may simply generate feelings and behavior (428).

and the size of the painting, and she has done so as a result of an alief (Brownstein and Madva 418).

Continuing this line of thought, perhaps if aliefs are capable of alleviating felt tension between us and our environment, they can also alleviate felt tension between our beliefs and our behavior. Brownstein and Madva explain that aliefs have affective and/or behavioral components (419). The affective component refers to the idea that, in response to a feature of our environment, aliefs induce a certain feeling in a person S, or cause her to hold certain attitudes towards this feature (Brownstein and Madva 425). For example, the museumgoer may feel "disoriented" in response to the painting (Brownstein and Madva 419). The behavioral component refers to the idea that in response to a feature of our environment, aliefs cause us to automatically behave in a certain way. For example, the museumgoer may move away from the painting upon feeling disoriented (Brownstein and Madva 419). What I seek to investigate now is the question "If an alief, in response to perceived tension, leads S to experience feelings and conduct behavior that seem to indicate that she actually believes p, are we justified in saying that she has deceived herself?"

I will again illustrate this possibility through Violet. When Violet purchases the fur coat, she experiences an inner tension between her belief in ¬pV and her action of purchasing the coat. Now what happens if Violet's aliefs lead her to automatically respond first by holding attitudes and then by behaving in ways that seem to suggest that she actually believes pV? For example, it seems plausible that Violet's aliefs might try to reduce her felt tension by altering her affect, e.g., by leading her to adopt a more relaxed attitude towards animal rights and fur. This affect, in turn, may then lead her to purchase leather boots or gloves lined with sheepskin—both behaviors which seem to indicate that Violet actually believes pV. If Violet's aliefs impact her affectively and/or behaviorally in such a manner, are we warranted in saying that she has successfully deceived herself? I would like to argue that aliefs do not have this power, but first, let us further examine what is happening when aliefs take control.

According to Brownstein and Madva, "Alief is a relation between an agent and 'F-T-B-A' content: feature-tension-behavior-alleviation" (420). The F-T-B-A sequence models how aliefs reduce the felt tension between the self and its environment, or for the purposes of our discussion, the self's beliefs and the self's actions. "Feature" refers to the component of the person's environment that turns an alief "on" (Brownstein and Madva 420) and tensions refer to "Automatic affective responses that are . . . geared towards immediate behavioral reactions" (Brownstein and Madva 421). "Behavior" is the physical or affective reaction to the felt tension, and "alleviation" is a sense of relief that occurs if the behavior successfully

eliminates the felt tension (Brownstein and Madva 422). Importantly, if the B-stage is completed but the subject does not experience relief, she repeats or modifies her behavior until she does (Brownstein and Madva 423-24). Thus, an alief may activate this series of automatic processes for the museumgoer: "Really big painting! (F) Do not have a good view! (T) Shift two inches back (B) . . . not good (Unalleviated) . . . shift two inches back (B) . . . almost there (UA) . . . shift one inch back (B) . . . perfect! (A)."[7]

If we run Violet's situation through the F-T-B-A formula, how might her aliefs act to alleviate tension? Modeling her F-T-B-A sequence after the museumgoer's, I propose that Violet's aliefs should affect her as follows: "I just purchased a fur coat! (F) Purchasing animal fur is wrong! (T) Immediately return coat to store! (B) Ah, now I feel better! (A)." Note that nowhere in this process does Violet seem to behave or exhibit affective or behavioral signs that indicate that she actually believes pV. Further, it seems much more likely that Violet's aliefs will result in this outcome, rather than behaviors or an attitude that would suggest that she actually believes pV. Returning the coat safeguards her self-image as a vegetarian and *fully* eliminates her felt tension, as she no longer has the coat to wear or is supporting the fur industry with her dollar.

Perhaps the reader is unconvinced that Violet's aliefs would lead her to respond to the tension by returning the fur coat, so let us suppose that Violet's aliefs *do* lead to affective and behavioral changes that seem to suggest that she believes pV, as outlined above. If her aliefs were to result in such changes, we may revise Violet's F-T-B-A sequence as follows: "I just purchased a fur coat! (F) I do not think humans should purchase animal fur! I'm going to purchase other animal-derived clothing products! (B)"[8] The "B" component here, as previously discussed, may also be accompanied by affective changes, such as increased feelings of acceptance towards the idea that it is not immoral for humans to purchase and wear fur. Lastly, if all goes well, Violet's tension will be alleviated . . . but will it?

I'd like to argue that it is highly suspect to claim that Violet will experience alleviation like the museumgoer would. Remember, Violet is a passionate vegetarian. So, if her aliefs were to lead her, however automatically, to form an attitude and then behave in a way that suggests she believes pV, it seems that she would have a sense that something still isn't right. To be clear, just because aliefs are relatively automatic does not

[7]F-T-B-A sequence partially modeled after R-A-B sequence (Brownstein and Madva 419).

[8]It's unlikely that Violet's aliefs would say exactly this, but to illustrate the sort of behavior they may lead to, the description of the B process is exaggerated.

mean that Violet is completely unaware of the behaviors and feelings they induce. So, if as a result of her aliefs Violet were to form attitudes and begin behaving in a way that aligns with pV, I argue that she would sense a disharmony between her actual belief in ¬pV and these attitudes and behaviors. Thus, because vegetarianism and animal advocacy are so integral to Violet's self-image, it is wrong to say that altering her behaviors and attitudes in a way that better aligns with a belief in pV will fully eliminate her tension. It is more consistent with her self-image to say that Violet's automatic response would be to donate the coat to a homeless shelter, burn the coat so that no person can ever purchase or wear the coat again, or, of course, simply return the coat! But behaving in a way that aligns with pV will not fully alleviate her felt tension, because these behaviors directly contradict a belief that, as previously claimed, she must be consciously aware of.

But perhaps Violet's case is an exception—maybe others will feel total relief by behaving and forming attitudes that align with the opposite of their original belief. Again, I reason that just as the museumgoer must make several behavioral adjustments before finding relief, a person S will always need to make further behavioral adjustments if her initial behavior aligns with p. That is, as a result of T, S's aliefs may lead her to hold attitudes and behave in a way that best aligns with a belief in p. Even if she senses it less deeply than Violet did, I'd still like to suggest that S will feel some discomfort in holding these attitudes and behaving in these ways while also being aware that she holds a belief against p. Thus, it seems that a "deceptive behavioral adjustment" does not terminate the F-T-B-A sequence: S could always find a more satisfying solution that more directly addresses her tension. In other words, if S's sequence were to result in beliefs and attitudes that align with p, we would again face a version of the problem that arose in my critique of Audi—namely, that S cannot be (at least on some level) consciously aware that she holds attitudes and is behaving in a way that aligns with p while also being aware of her belief in ¬p. To be consciously aware of both would only intensify and prolong S's felt tension between action and belief. I then conclude that the fact that S's tension is not fully alleviated indicates that S does not actually believe p, and we are again not justified in saying that S has successfully deceived herself.[9]

[9]It is possible that S may be "partially deceived," but whether this is even possible is a topic to be explored in a future essay.

Brownstein and Madva's discussion also accords with my argument that we cannot make claims of self-deception based off of behavior. Brownstein and Madva explain that in most cases, "The truth-taking view, which attributes belief on the basis of agents' reflective judgments and avowals, outperforms the alternatives," (415) a major alternative being the one posed in the previous paragraph that "attributes beliefs based on the agents' spontaneous actions and emotions" (415). This is to say that S's beliefs are simply what, on reflection, she finds that she believes—we should not infer that S believes one proposition or another from her behavior alone (Brownstein and Madva 415). That is, if Violet were to reflect on her act of purchasing the coat, she would most likely agree that this act went against her morals. She would come to understand that, through behaving in a way that distracted her from acknowledging her true belief, she was trying to convince herself that her actions were not immoral.

Ultimately, I conclude that behavior is not evidence enough to show that Violet has actually deceived herself (i.e., believes pV). Upon reflection, Violet is almost sure to realize that her behaviors were an automatic response initiated to prevent her self-image from becoming fractured. Particularly, the only reason these behaviors were initiated in the first place was to relieve the tension between Violet's actions and her belief in ¬pV. In absence of a felt tension, there is no reason to think that Violet ever would have behaved in such a way and would have continued to assert that she believes that humans should not purchase or wear animal fur. Hence, we are not justified in saying that her behaviors prove that she is deceived.

A Justification for the Practice of Attempted Self-Deception

At this point, it is unclear that self-deception is an attainable mental state. When Gergen makes this conclusion, he goes on to discuss what he thinks are various social uses of the folk concept of self-deception. On a somewhat different note, I would like to explore potential epistemic uses of self-deception, and in particular argue that S's *attempting* to deceive herself enables her to become clearer on her beliefs.[10] Let me explain what I have in mind.

In my discussion of aliefs and beliefs, I noted something of significance: unlike the museumgoer, who eventually settles in the right spot in front of the painting, S can never, through a process of affect and

[10]I use "attempting" here because on my analysis it is unclear that S can ever be successful in her deception.

behavior alterations, "settle" in a state of alleviation. Based on this concept, I would like to propose a model of (what we think of as) self-deception that is more active than those previously discussed. What I propose is that, upon *consciously* noticing some tension between her belief in ¬p and her actions, S attempts to convince herself that she believes p, and offers herself reasons in support of p to strengthen her affirmation. She hopes that through this process, she will find her reasons in favor of p so convincing that she "forgets" that she ever believed ¬p. What I suggest here is that in most cases, S will fail in this endeavor—she will rarely[11] be convinced by her reasons in favor of p. This may seem frustrating, but it is actually quite helpful. To see what I mean, let's again turn to Violet.

When Violet purchases the fur coat, she is, I have argued, conscious of her belief in ¬pV, but she is uncomfortable in the fact that her act of purchasing the coat is in tension with her belief in ¬pV. Violet is then on the defense; she convinces herself of pV and offers herself reasons in support of this claim. She may reason that because she purchased the coat at a thrift store, she is not directly supporting the fur industry. Or, she may claim that the fox is already dead, so it is not as if her act of purchasing the coat is actually hurting any animals.

What I suggest will then happen is that the more Violet attempts to justify her decision to purchase the coat, the more she realizes that her reasons are not strong enough to convince herself that her purchase is justified. And the more she attempts to provide reasons that her purchase *is* justified, the more difficult it becomes to forget about her original belief that purchasing the coat is immoral. That is, it is true that through shopping at a thrift store Violet is not directly supporting the fur industry, but the store is still benefiting from its selling of fur. Or she may see that, just because her actions have not lead to the death of any animals, an animal still had to give up its life just for her to stay warm.

If Violet does find she can respond in such a way to her reasons in favor of pV, she comes to realize that she does not actually agree with pV. What seems to happen then is that Violet's belief that humans should not purchase or wear animal fur is only reaffirmed when she attempts to justify an avowal that her purchasing of the coat was not immoral.

This, I suggest, is the epistemic value of an attempted self-deception. When a person S attempts to deceives herself about a belief in ¬p, it is because she has committed an act that is in tension with this belief. Realizing this, she then attempts to convince herself that she believes p. But

[11]Counterexamples will be offered shortly.

the more she travels down this rabbit hole of an attempted self-deception, the more she realizes how weak her reasons in favor of p are compared to her belief in ¬p. And her reasons in favor of p *should* prove to be weak—she is trying, after all, to affirm the opposite of a strong belief she had in the first place.[12] Again, when we attempt to self-deceive we do not truly believe p: the only reason we try to convince ourselves that we believe p in the first place is to alleviate the feeling of tension between our action that went against ¬p and our belief in ¬p.

Thus, because our reasons in favor of p often prove to be weak, I argue that we should allow ourselves the attempt to self-deceive in order to potentially realize that we should not have been striving to convincing ourselves of p all along. Further, when we allow ourselves the attempt to believe p, we end up reaffirming our belief in ¬p as we often find that ultimately our reasons in support of p are unconvincing.

Objections and Concluding Remarks

The reader may by now have several questions about my characterization of self-deception, which I seek to address. First, the reader may object that self-deception is not the only way for S to affirm her belief in ¬p. For example, Violet might instead adopt a reflective approach: upon registering her felt tension and reflecting on it, she finds that she feels uncomfortable precisely because she strongly believes ¬pV. In this way, she is also able to "be in touch" with her beliefs. But the reason that I am arguing for attempted self-deception as a means of affirming one's beliefs is that it is often the natural human defense for committing an action that goes against our beliefs, and many of us are not always so willing to acknowledge that we have committed such an action. Further, it seems that when we allow ourselves the attempt to self-deceive, we become clearer and more accurate about what we believe than we would be simply by reflecting on the tension. Specifically, in the attempted self-deception S unintentionally allows herself the opportunity to "play devil's advocate," if you will. She attempts to see things from a different perspective, and when she does, she realizes that she doesn't actually agree with this perspective.

Building on this last idea, the reader may object that it is possible that, in her attempted self-deception, S finds that she is convinced by the reasons she offers in favor of p. Perhaps Violet actually is convinced by her

[12]We know that S's belief in ¬p is strong, because otherwise she wouldn't feel so uncomfortable in the fact that she has committed an act that goes against this belief.

reason that she is not directly supporting the fur industry. It is true that the thrift store is benefiting from selling the coat to Violet, but the thrift store is a charitable organization—ultimately Violet's money is being put to good use. Similarly, she may be convinced that she did not technically harm any animals through her purchase—the store received the coat by donation. And suppose that Violet finds that she is not unconvinced by any of her reasons in favor of pV. Then are we warranted in saying that Violet's deception was successful?

Again, I do not think that we are. It seems that part of our notion of self-deception is that S attempts to convince herself of the opposite of a proposition which she really does believe. And if ¬p is something that she truly does believe, she should not be easily convinced by any of her reasons offered in favor of p. That is, the attempted self-deception should result in a reaffirmation of ¬p. What seems to be happening to Violet as described above is that she is not deceiving herself, but rather, through examining her arguments in favor of both pV and ¬pV, finds that pV better captures her overall self-image. While Violet is a passionate vegetarian, part of her self-image is the desire to maximize the good in the world through her actions. If slightly compromising on her vegetarianism will allow her to contribute to a charitable organization, perhaps the belief "It is not immoral for humans to purchase or wear fur" better captures what Violet believes. Thus, I propose that in "self-deception" as we speak of it, S's belief in ¬p is generally so firm as to not be easily uprooted by reasons in favor of p. If ¬p is easily uprooted, then we no longer have an instance of S attempting to self-deceive but rather an instance of S changing her mind or becoming clearer on what she believes. In any case, both possibilities carry epistemic weight: S emerges better understanding what she believes.

We see a similar sort of idea in Chapter 1 of Jennifer Church's *Double Consciousness in Everyday Life*. Church explains that "At their best [quarrels] are supposed to reveal new facts and values—facts and values previously overlooked by one of the quarrelling parties, or facts and values to which both parties were blind" (8). If we think of self-deception as an inner quarrel—attempting to believe p when we actually believe ¬p—we see that the attempted deception may actually help us to better understand all the reasons why we believed p in the first place, reasons that we may have been previously unaware of. Therein lies the epistemic value of an attempted self-deception: by considering her alternatives and finding them unsettling, S develops a more complete understanding of why she believes ¬p. Further, if the attempted self-deception transitions into S's changing her mind, S was still able to survey her reasons and better understand why she believes what she believes.

The reader might hope that I concede that self-deception need not always be a positive practice, but as I have defined it, I still argue that it will be. For instance, we might imagine a student, let's call her Messy Margaret, whose dorm room is in a constant state of disarray. Margaret unconsciously knows that it is *not* the best idea to leave mountains of laundry scattered on her carpet and allow her trash bin to overflow, yet deceives herself into thinking that her messiness is justifiable as a result of her busy schedule. One concern is that even if Margaret does, as a result of reflection, find that she actually believes that her messiness is *not* justifiable, this does not mean that she needs to act on this belief and become a neater person. In response, I point out that Margaret's failure to act on her newly realized belief is *not* the fault of self-deception, a process which only led her to understand what she truly believes. Instead, her failure to align her beliefs and her behavior is a matter of willpower.

In an alternate scenario, we might imagine that Margaret simply chooses to *ignore* any felt tension between her messy actions and her underlying belief that she should be tidier, and instead mindlessly asserts to herself that she is fine with her messiness. In this case, Margaret is not even able to gain the epistemic clarity that, as I have argued, self-deception produces. Does this suggest that self-deception is only of use to the reflective, thoughtful subject? Surely not. Self-deception is itself an inherently reflective process as I have described it—in order to "successfully" deceive oneself, one must provide convincing reasons in favor of p. Again, this should result in S's either a) realizing that these reasons are not as persuasive as she originally thought or b) finding that she actually does believe p and revising her beliefs. Hence it is possible for S to disregard any felt tension and insist that she believes p all without the epistemic clarity whose benefits I have preached—but fervent insistence is not deception. To say that S truly believes p requires more than that she fiercely claims to believe p; S must have strong *reasons* pointing in favor of p. Thus, blind allegiance to certain beliefs may surely produce negative behavior, but the process of self-deception, I claim, is always beneficial to the subject. Similarly, self-deception may reveal to S that her beliefs lead to morally questionable actions or behavior, but again, it is the product, and not the revelatory *process*, that if anything, is negative.

In this paper, I explored several ways in which the mental state of self-deception may be possible. I found that neither Audi's model nor automatic processes (aliefs) are capable of producing this state, and concluded that it is unlikely that one can actually "deceive herself." I then offered my own thoughts on how *attempted* self deception could serve the individual, and argued that self-deception is a useful tool in that it allows S to travel down a path which assists her in becoming clearer on what she truly believes.

Works Cited

Audi, Robert. "Self-Deception and Rationality." *Self-Deception and Self-Understanding*. Ed. Mike W. Martin. Lawrence: UP of Kansas, 1985. 169–74.

Brownstein, Michael and Alex Madva. "The Normativity of Automaticity." *Mind & Language* 27, no. 4 (2012): 357–493.

Church, Jennifer. *Double Consciousness in Everyday Life*. Manuscript–Chapter 1, 2017.

Gergen, Kenneth. "The Ethnopsychology of Self-Deception." *Self-Deception and Self-Understanding*. Ed. Mike W. Martin. Lawrence: UP of Kansas, 1985. 228–43.