Demystifying the Deep Self View

Abstract: Deep Self views of moral responsibility have been criticized for positing mysterious concepts, making nearly paradoxical claims about ownership of one's mental states, and promoting self-deceptive moral evasion. I defend Deep Self views from these pervasive forms of skepticism by arguing that some criticism is hasty and stems from epistemic injustice regarding testimony of experiences of alienation, while other criticism targets contingent features of Deep Self views that ought to be abandoned. To aid in this project, I provide original naturalistic analyses of "Self" and "internality" that replace the view's metaphorical language with common-sensical concepts that make clear their usefulness.

1 Introduction

Folk explanations of exemptions from blame due to conditions like compulsion hold that in these cases a person's "brain makes them do it." There seems to be something to this idea, although if taken quite literally, it might seem that our brains make us do all that we do. It is in part this very fact that has been thought to make mysterious how anyone could be responsible for anything. Deep Self views of moral responsibility represent a promising way forward in the quest to show how actions that are caused by our brains can nevertheless express our agential perspectives, while also making sense of this folk notion of exemption.[1] These views hold that hold that what makes an agent an appropriate target for praise or blame on the basis of her action is that it issues from some privileged subset of her psychology that separates the agential from the random firings of her brain. By contrast, when an agent is exempt due to something like compulsion what exempts her is that her action issues from some noisy neural process that stands outside of the agential part, whose boundaries the Deep Self theorist aims to sketch.

Deep Self views, however, currently occupy a precarious position in the literature. Though there have been no dearth of defenders, misunderstanding about what precisely the commitments of Deep Self views are has led to an air of heavy suspicion by critics and those unacquainted with the Deep Self tradition alike. For one, Deep Self are sometimes thought of as positing the existence of something quite mysterious. This is unfortunate, as these views arguably require much less fanciful metaphysics than many competing views, and should be considered some of the most naturalistic, empirically grounded views of moral responsibility on offer. It is not hard to see how the

[1] Deep Self views are sometimes referred to as "True Self" "Real Self" "Self-Disclosure" "Identificationist" or "Attributionist" views. See Wolf (1987, 1990) for the initial title and characterization of the set of views as "Deep Self" views.

misunderstanding might have taken place. Deep Self theorists tend to articulate their views somewhat impressionistically with semi-metaphorical words and phrases like "speaking for," "internality," and "alienation." Furthermore, in reading Deep Self accounts it can seem its defenders have set for themselves an impossible task: to locate some central most authentic seat of pure unconflicted agency amidst the swirl of an imperfect human's actual mental states. Not only might this seem like a fool's errand, but it also might seem morally questionable to hope that we might be able to identify ourselves with some pure and stable Self, distancing ourselves from the consequences of our everyday actions caused by tempting, unwise, and conflicting urges.

But, as I will argue, the advantages of Deep Self views of moral responsibility needn't be tethered to any such lofty ambitions or metaphorical language. Elsewhere I have argued for a particular version of a Deep Self view[2], but my ambitions here will be broader: to vindicate Deep Self views from these common causes of suspicion. In doing so, I will aim not only to dispel misconceptions about Deep Self views, but also to turn a critical eye to certain features of leading Deep Self views that I believe have fed into the criticism of the Deep Self project at large. My plan in the rest of this article is as follows: in §2 I highlight the benefits of Deep Self views, in §3 I identify and diffuse the major sources of skepticism; in §4 I give common-sense naturalistic analyses of the Self and internality, two of the most mysterious sounding concepts in the Deep Self lexicon; and in §5 I briefly conclude with suggestions for Deep Self theorists and critics going forward.

2 Common-Sense Motivations for Deep Self Views

Despite the heavy air of suspicion surrounding motivations for Deep Self views, two of the strongest points in its favor come from its ability to provide common-sense explanations. First, it is able to show why exemptions from responsibility have nothing to do with what the agent in question *would have* done in a counterfactual world and only with how it came to be that she has done what she in fact has done in the actual world. Second, it promises to give a naturalist account of the phenomenology of non-ownership of one's action which is shared by many people with certain kinds of psychological and neurological disabilities.

2.1 Using Patterns of Actual Mental States as Criterion for Responsibility

Deep Self views look just at the actual mental states that lead an agent to act in the way that she in fact does on some occasion, and provide a decision procedure for identifying whether this profile of mental states are of the right kind. If they are, we can properly say that they issued from the agent, and she is now, in principle, be a target of praise or blame on the basis of how she acted. This decision procedure does not rely on having to think

[2] [Redacted]

about how, under what circumstances, or if, the agent might have acted differently than she in fact did.

This advantage has tended to be pitched as primarily dialectical. In focusing on actual mental states of the agent, the Deep Self theorist avoids thorny questions about the relevance of whether the agent could have acted otherwise, and about whether an agent's sensitivity to reasons can matter if she doesn't respond to the reasons in the actual world.

Less well appreciated is the fact that powerful common-sense intuitions tell in favor of the idea that the explanation for why you in fact did what you did contains all of the factors that make you responsible or not. It would seem quite inappropriate for an agent to attempt to absolve herself of responsibility for some action by pointing to factors that were not in any way explanatory of why she acted in such a way.[3] This is, arguably, the most important lesson of so-called Frankfurt cases.[4] In these cases an agent is going to perform some action, φing, for which we all agree she would be intuitively morally responsible were she to go through with it. Unbeknownst to her, though, there is someone waiting in the wings who would be activating a chip in her brain causing her to φ if she did not independently go through with φing. But, as it happens, she goes through with φing of her own volition and no intervention takes place.[5] These sorts of cases have spawned a large literature about whether or not any particular articulation of the case can decisively establish a foolproof example of an agent who has no "alternative possibilities" in the specific senses used invoked in various theories of moral responsibility, but is nevertheless responsible for her action, thus providing a counterexample to the view.[6] Independently of the outcome of those debates, though, Frankfurt cases remind us that it seems much less important to figure out whether or not there are possible worlds in which the agent does not φ than it does to figure out why she actually φs.[7]

---

[3] As Carolina Sartorio puts it, "if a factor is completely irrelevant to why you acted, it seems that it cannot be used to excuse your behavior" Sartorio (2016): 2. See also Mele (2006).

[4] See also McKenna (2008) for a similar take on the relevance of Frankfurt-cases.

[5] Frankfurt (1969).

[6] Debate rages on about, for example, whether or not it is methodologically appropriate to make the assumption that the intervener knows what the agent will do before she does it, or whether or not the action the intervener would cause would be identical to agent's actual act. See Fischer (2010) for an overview of the current state of the literature in which he compares it to the state of the literature on Gettier cases.

[7] This is consonant with Frankfurt's own articulation of the moral of his cases:

The fact that a person could not have avoided doing something is a sufficient condition of his having done it. But…this fact may play no role whatsoever in the explanation of why he did it. It may not figure at all among the circumstances that actually brought it about that he did what he did, so that his action is to be accounted for on another basis entirely. Even though the person was unable to do otherwise, that is to say, it may not be the case that he acted as he did *because* he could not have done otherwise (Frankfurt [1988]: 8).

Consider also someone who loves doing something so much that she would never consider abandoning it who also develops a compulsion towards engaging in it. Surely we should be able to continue to blame and/or praise her for engaging in the activity even if, were she to attempt to abandon it, which she would never want to do, she would be compelled to continue.[8] If this common-sense intuition is right, it seems that we'll have to have some way of differentiating between compulsive and non-compulsive mental states without looking at whether the agent could have acted otherwise.

2.2 Explaining Alienation

Another of the attractions of Deep Self views is that they offer an elegant explanation for an otherwise puzzling psychological phenomenon. People sometimes experience some of their own behaviors as being divorced from their ordinary agency such that their own behavior feels quite alien. This experience is especially familiar to people who have certain kinds of compulsive disorders, tics, or conditions caused by neural cross-wiring.

When people experience this kind of puzzling alienation from their own behavior, they sometimes liken it to another being temporarily taking control of them. For example, Josh Hanagarne, in his autobiographical novel, *The World's Strongest Librarian*, writes: "I saw Tourette's as a separate being; a parasite that I was in a relationship with. I named her Misty, short for 'Miss T'".[9] Blogger Abi Flynn describes her experience with misophonia, a neurological condition in which hearing harmless repetitive sounds causes extreme distress and the urge to lash out as involving the feeling of being possessed. She writes,

> I remember it felt like being possessed, like it was not me being so offended by the sound, but a sudden demon inside me that would arise on cue and rage around causing me intense suffering and dis-ease.[10]

While these contemporary writers merely attempt to make sense of and describe the feeling of non-ownership of their actions in terms of possession, it is probably no coincidence that some of these conditions, like Tourette's syndrome, were once thought to *actually* be a form of demonic possession.[11]

The Deep Self approach promises to explain how an agent could, in a less mysterious way, fail to be identified with her action, and why we might really be mistaken to hold the agent responsible for what she does in the normal sense in cases in which she is alienated in this way. The idea is that the motivation that leads a person to ultimately act usually has some sorts of further consonance with her psychology such that it makes sense to say that her resultant action properly issues from her agency. In cases of tics,

---

[8] See also Frankfurt's 'Willing Addict' and Sripada's 'Willing Exploiter' (Frankfurt [1971], Sripada [2017]: 802-803.)

[9] Hanagarne [2014], pg. 65.

[10] Flynn [2017]. For more on misophonia, see Braut et. al (2018) and Kumar et. al (2017). [Redacted].

[11] Germiniani et. al [2012].

brute compulsion, and misophonic outbursts, the agent is moved to do something against her will by a rogue urge that overpowers her normal agential process. There are different views of just what the will, or normal process of agency, amounts so. Different Deep Self theorists propose different demarcations within agential psychology to explain which subset of the agent's motivational states can speak for the agent. The most popular views argue that it is only the motivational states that mesh with what the agent values, plans, cares about, or endorses.[12] I will refer to these further mental states with which motivations are meant to mesh with as "deep self mental states."[13]

So, while there's no demon or anyone else responsible for an action that an agent is alienated from in this sense, when it fails to issue from agential mental states the agent isn't responsible for it either because it's not really an expression of *her* in the normal sense. Oftentimes these feelings are tracking a real lack of agential input in the output of what one ends up doing. While there's no one else arising within the agent and taking over, her feelings that her agential self is not playing its usual role in producing her action are veridical.[14]

3 Diagnosing the Skepticism

3.1 The "*Deep*" in "Deep Self"

One of the sources of skepticism about Deep Self views is suspicion about the concept of a "deep" self. It is, in a way, unfortunate that the "Deep Self" name is the one that has

[12] See, for example, Frankfurt (1971, 1987, 1992 2006), Watson (1975), Mitchell-Yellin (2014, 2015), Bratman (2003), Shoemaker (2003, 2015a, 2015b), and Sripada (2016). Views that put forth other candidate deep self mental states include Susan Wolf's "sane Deep Self view" on which deep self mental states must meet further "sanity" requirements (Wolf [1987]); David Velleman's view, on which deep self mental states are desires to act in accordance with reasons (Velleman [1992]); and coherentist views on which deep self mental states are those that bear special relationships to the agent's other mental states either by being relatively unopposed by other states (Arpaly and Schroeder[1999]) or by being narratively coherent (Matheson [2018]).)

[13] What does it mean for a special mental state to "mesh" with one's motivation? Many theorists seem to think about the relation as being causal: an agent is responsible for φ-ing if the motivational state that causes the agent to φ is itself caused in part by the agent's deep self mental states. Responding to a challenge that causal dependence is insufficient for expression of deep self mental states (Levy 2011) Sripada and Shoemaker put forth a "content harmony" relation (Sripada [2016], Shoemaker [2012, 2015b]). On this view, an agent is responsible for her action only if the motivational state that she acts on is congruent with the content of the deep self mental state in some sense.

[14] On my own view, this is only part of the story.. An agent may be alienated from a particular mental urge, while nevertheless retaining responsibility for the management of that urge, thus retaining at least partial responsibility for acting in accordance with it. (See [redacted]. I set aside these complications here to focus on the initial motivation for thinking such actions call out for an explanation as to why agents experience some form of alienation from them.

stuck, as it tends to evoke thoughts of a quite ambitious project to locate a central, fundamental, all-important seat of agency within the sea of an agent's mental states. Many people, myself included, find it implausible to think that any such pure seat of agency exists. But this is no indictment of the Deep Self project in general, as the aims of a Deep Self theorist in practice can be much more modest.

There is no consensus among Deep Self theorists about just what commitments are taken on by adopting the language of the "deep" self. For example, David Shoemaker writes that

> the 'deep' in 'deep self' simply refers to the psychic element's place in an agential structure as the ultimate psychological *source* of various 'surface' attitudes subject to its governance.[15]

So, for example, cares are deeper than ordinary first-order desires since, for example, caring about your family is the *source* of a desire to take your daughter to soccer practice. The meaning of "deep" here does not imply any sort of strong metaphysical commitment to the Self. On the other end of the spectrum, Chandra Sripada thinks talk of deep selves commits him to the existence of "fundamental conative states that robustly and globally shape action," the existence of which he takes to be a substantive claim about actual human psychology.[16] Deep selves, for Sripada, are presumably 'deep' because on his view they play a crucial role in helping to explain a wide variety of agential phenomena including but not limited to: moral responsibility, normative reasons for action, happiness, and weakness of will. A similar but, in theory, distinct idea is the thought that all of an agent's mental states of the kind that are proposed to play the role of deep self mental states together form some sort of whole which either constitutes or provides us with some particularly important insight into the agent's Self.

I take these latter two conceptions to be merely contingent features of the set of views generally recognized to belong to the Deep Self family of views. Each view does need some story to tell about what privileges actions that relate to deep self mental states such that they are the ones on the basis of which we are permitted to hold an agent responsible. However, the versions of this story on which the deep self mental states together play a foundational role in the core of an agent's conative personality or are together constitutive of the agent's Self only represent a couple of the options for fleshing out this story, among many other possibilities.

---

[15] Shoemaker (2015b): 43.

[16] Sripada (Unpublished Manuscript): 15. While it does not seem to me that any such broad sweeping claims about human psychology are required for proponents of Deep Self, Sripada thinks there is much less cause for empirically-driven skepticism about the existence of such deep selves than what many philosophers have been led to believe. According to Sripada, while certain segments of social psychology have been very influential as a source of data for philosophers, data from neuroscience, human behavioral genetics, and personality psychology is all fairly friendly to the idea of robust deep self psychology.

Further complicating these issues, as Lippert-Rasmussen points out, people tend to conflate two different connotations of the phrase "deep self." On one conception, deep self mental states have special *authority* for the agent, and on another, deep self mental states have more to do with *authenticity*. As he puts it, on authenticity conceptions of the Deep Self, a person's Deep Self

> … is the person's deepest and most genuine commitments and desires…deep, idiosyncratic longings and repressed desires are strong candidates [for deep self mental states] on [this] account.[17] (20).

This conception of the deep self is, to my mind, not relevant to questions of moral responsibility.

Confusion in the literature between the two senses of "Deep Self" is presumably part of what leads Nomy Arpaly to her particularly damning accusation of Deep Self views. She has us imagine a woman, Lynn, who discovers she is a lesbian but would much rather have not come to such a discovery and does not want to be motivated by such desires. Arpaly continues,

> If Lynn were to go to her favorite college professor for help, she would likely be told that she should try to accept herself for who she is, refrain from attempts to suppress her true self, and so on. If, on the other hand, she were to read the moral psychology literature and believe its claims, she would probably conclude that she was right and her homosexual desires are not truly her own.[18]

But it need not be any part of a Deep Self view to hold that Lynne's lesbian desires are not an authentic part of who she is or that she should resist them. A Deep Self view should merely say that if she were to engage in a sexual act with a woman without in some sense valuing/ endorsing/ planning on/ caring about doing so, her action would be compulsive or lacking in agential authorization in such a way that would undermine her being an apt candidate for moral responsibility. It is perfectly consistent to additionally hold that Lynne ought to embrace her lesbian desires as being an authentic part of her identity. Deep Self theorists ought to be clearer in rejecting the relevance of the authenticity conception of the Deep Self and instead understand deep self mental states as those that have the authority to speak for the agent; the mesh of deep self mental states with effective motivation needn't be understood to be anything over and above a condition for ownership over one's action in the sense relevant for moral responsibility.

Although I've highlighted that the aims of Deep Self theorists do vary quite a bit in terms of how modest they are, I hope I've shown that worries about ambitious

[17] Lippert-Rasmussen (2003): 20.

[18] Arpaly (2002), 16.

psychological theorizing shouldn't be taken to be any kind of knockdown argument against the Deep Self project as a whole.

3.2 Doubts About the Credibility of Testimony about Alienation

When advocating for a particular candidate deep self mental state, theorists generally try to argue that they have identified a mental state kind that bears the mark of internality. Internality is usually understood as being a property of a mental state kind such that for any tokens of that kind a person cannot be alienated from them. This invites the reading that what's special about deep self mental states is just that people aren't inclined to disavow them; their feelings or professed feelings of alienation do a lot of heavy-lifting. The idea that we should put a lot of weight on the testimony of people who claim to feel alienated from their motivations is surely a major source of skepticism about Deep Self views

But, as Agnieszka Jaworska points out, we can distinguish internality in an ontological sense from subjective active identification that is based on whether the agent perceives aspects of her psychology as being her own.[19] While there is a possible view on which the ontological category of internal mental states with which an agent can rightly be identified amounts to nothing more than the states with which the agent takes herself to be identified with, such a view would require an argument. These senses of internality are not wholly unrelated, however, as non-self-deceptive subjective identification provides us with defeasible evidence of internality. So even if we have the ontological sense of internality in mind, we might not have good reason to believe in its existence without putting some stock in testimony about feelings of alienation. There are several reasons that people doubt that testimony about alienation provides good evidence of the existence of mental states from which a person is actually alienated.

First, the claim that certain of your mental states are *really* yours and that some are somehow not can seem nearly paradoxical. If they're not your mental states, whose are they exactly? And if you're the one being motivated by them, how could they not be *your* motivations? Richard Moran is among the recent outspoken skeptics of this way of talking, finding it ultimately incoherent. If a desire really belonged to no one, he argues, it would not exist at all. In fact, he thinks, refusing to identify with a desire requires you seeing it as *yours* to identify with or not in the first place. One of the mistakes made by Deep Self theorists, according to Moran, is that they think that in considering one's own approval of a desire we would need to identify a particular person the desire belongs to at all. On the contrary, it should be sufficient for it being yours that you are the one who apprehends its attractions.[20]

[19] Jaworska (2015): 531.

[20] Moran, forthcoming.

It's true that we're used to talking about mental states not being ours only in cases when they are someone else's, which is probably one of the reasons that people have been inclined to talk about demon possession etc. in cases of alienated urges. But, as Frankfurt notes, it is actually not at all obvious, except in a fairly trivial sense, that all of our desires belong to us since they do not belong to anyone else. We only attribute some of the events in the history of a person's body to that person in a strict sense; some of them are mere happenings, such as getting lurched forward on a bus or experiencing a bodily twitch. It's not so much that considering a movement as one's own is in general a feature of acting to move part of one's own body, but rather that there is something phenomenologically distinctive from the case of a mere twitch. Just as it would be unfair to say that because such behavior is not attributable to anyone else, it must be attributable to the agent, so too is it unfair to make this inference in regards to desires. Of course this does not decisively prove that one may be alienated from one's own desire in a sense that makes sense to talk about as that desire not truly belonging to that person, but the evidence in favor of this is not dissimilar to the kinds of evidence we have of bodily-alienation.[21]

But unlike bodily twitches, suspicion about the very idea of being alienated from one's own mental states is often coupled with the concern that feelings of alienation are evasive, fabricated, illusory, or the result of self-deception. This is a second source of doubt regarding testimony about alienation. Terence Penelhum expresses this line of criticism particularly forcefully. He says, regarding an agent's expression of the fact that his motivating desire does not truly belong to him, that it is a

> form of moral trickery… involv[ing] an extension of the notion of non-identification with one's own desires and behavior from the level of harmless and even mildly illuminating metaphor to that of gross literal false-hood. To say harmlessly that one is governed by a desire that is not one's own is to utter a metaphor the literal translation of which is that one is governed by a desire that one does not want to be governed by. To say that the desire is not one's own and mean this literally is to say something obviously false: for the desire is operative and therefore exists, and is not someone else's. This obvious falsehood can be given the appearance of respectability with the aid of philosophical theories about the division of the soul's faculties; and it is a falsehood we are sometimes willing to swallow about others as well as about ourselves, as in the Gallic concept of the *crime passionel*. But we all know better.[22]

Arpaly and Schroeder's diagnosis is somewhat different:

> When people find that they have not been as rational, sane, prudent or moral as expected, they may experience…the cause of their misbehavior as an alien

[21] Frankfurt (1988): 61-62.

[22] Penelhum (1971): 670.

> intrusion[23]….In a culture such as our own, glorifying decisiveness, self-control and 'follow-through,' and with a tendency to medicalize failures of such traits, many agents will instinctively reject evidence of themselves as straightforwardly akratic, as having simply chosen poorly when they knew better. Instead, in some (and, it seems, a growing number) of circumstances, they experience their failure as apparently incomprehensible, an ugly intrusion upon their lives, and the psychological cause of this failure seems an unpleasant intruder.[24]

But again, allowing that a person can disclaim certain motivations as external is only as much of an opportunity for moral evasion or self-deceit as allowing that a person can disclaim certain movements of her body as external is. And yet, we do not regularly take this as reason to be skeptical of bodily twitches. What explains the difference here?

Like bodily twitches, the frequency of motivational alienation varies across individuals. I think that being moved to action by an external urge might be more common than is ordinarily recognized by skeptics. For example, I think that scratching an itch can sometimes be motivated by a desire that stands outside of a person's agency.[25] But, as this example shows, many of the situations in which the average person is genuinely alienated from a motivation are fairly trivial and/or commonplace. These cases may not even register as alienation since they don't occasion much self-reflection, making it so that recognizable cases of alienation from one's own desire are fairly rare. If this is right, one potential explanation as to why many feel that speaking of external desires would be tantamount to making up an excuse for one's action is that these people over-extrapolate from their own experience. When given a description of an agent acting in a way that she describes as alien, the inference to the best explanation might seem to be that she is making an excuse for her behavior if it is the case that if *you* were to speak about a similar action in such a way, you *would* be merely making an excuse for your behavior.

This problem is exacerbated by two factors. First, while overreliance on one's own experience in cases in which more deference is called for is widespread in general, it is an especially common occurrence in situations in which the person whose experience is owed some amount of deference is subject to epistemic injustice. And, indeed, people who frequently experience alienated motivational states are subject to the pernicious confluence of two significant epistemic injustices (factors that wrong them in their capacity as knowers). They are subject to testimonial injustice, credibility deficits on the basis of prejudicial characterizations of people with mental health disabilities. People with mental

---

[23] See also Buss (2012). Buss takes it that when some people speak of alienation, what they really mean is that they act with a lack of a willing attitude. But, as she puts it, "just as autonomous agency is compatible with stupidity and thoughtlessness, so too it is compatible with ambivalence, regret, disappointment, frustration, and self-criticism" (pg 655).

[24] Schroeder and Arpaly (1999): 383.

[25] I develop this point further in §4.1.

health disabilities are devalued as a group by society and are treated due to their identity group-membership as emotionally unstable, cognitively unreliable, or bizarre, as well as dangerous and morally suspicious: features that interfere with their perceived epistemic credibility.[26] These factors operate subconsciously, making it so that even the most sympathetic reinterpretations of testimony about alienation can easily mistake a projection of what they imagine the person is feeling that disregards the person's actual testimony for a charitable interpretation of it.[27] The disposition to disbelieve testimony about alienation is further activated by the fact that this testimony can sound, frankly, quite bizarre, especially when it invokes ideas of demon possession and the like. But these strange ways of characterizing the experience may stem in part from a lack of *adequate* vocabulary to describe it due to an unjust flaw in shared hermeneutic resources. People with mental health disabilities have long been systemically excluded from the institutions that seek to explain and make sense of psychiatric phenomena, which could very well lead to a failure to develop stigma-free ways of describing the experience.[28] Professional philosophy is not exempt as a potential contributing institution. Since, as Abigail Gosselin put it, reasoning capacity "is the currency of power, authority, and privilege" in the discipline, philosophers are especially vulnerable to the harms of self-disclosing any kind of psychological difference that may have the effect of undermining perceptions of their reasoning capacities.[29] When the people who experience a certain psychological phenomenon are disenfranchised from shaping the standard models of agential architecture, we should not be too surprised that their professed experience fits uncomfortably within them.[30] It is of course true that we should not take the testimony of people who claim to experience some psychological phenomenon such as motivational alienation as an unimpeachable fact, but I think we have at least as much reason to worry about our tendency to dismiss it.

That said, the problem of Deep Self skeptics over-extrapolating from their own experience to reinterpret experiences of alienation is also exacerbated by the fact that, until recently, most Deep Self theorists were not careful to distinguish compulsion from more ordinary agential phenomena such as weakness of will and carelessness. These early views tend to conflate the relatively minimal criterion for actions to count as attributable to agents for the purposes of agential appraisal on their basis with some more robust

---

[26] Jackson et al (2009): 167–168, Carel and Kidd (2014): 529, Kurs and Grinshpoon (2018).

[27] I take this point from Jackson (2017) who draws on Max Scheler's account of sympathy to describe a similar epistemic arrogance displayed in regard to reinterpretations of first-personal accounts of clinical depression. According to Scheler, sympathy "invariably pre-supposes what it is attempting to deduce" (Scheler 2008: 5). As Jackson is surely right to note, people who embody these problematic attitudes might themselves be part of the same stigmatized group and be subject to internalized stigma.

[28] For further discussion of this point, see Kurs and Grinshpoon (2018).

[29] Gosselin 2019.

[30] For another example, see Calhoun (2008) for discussion of the gendered dimensions of philosophy of action's failure to adequately account for clinical depression on models of agency. [Redacted].

conception of self-governing action or full-blooded "agency *par excellance.*" If any action that conflicts with one's deeply held values, agential plans, or endorsed course of action is spoken of in terms of agential alienation, it's not hard to see how a Deep Self skeptic could think "Well if *that's* what alienation means…that should not be seen as genuinely exculpatory. After all, I give in to temptation all the time, and I ought to be harshly judged for it!" Deep Self theorists who want to avoid this reaction ought to follow Arpaly and Schroeder (1999), Shoemaker (2003), Sripada (2017), Matheson (2017), Gorman (2019, Forthcoming) in offering views on which weak-willed actions are not automatically deemed actions from which the agent is alienated.

4 Providing an Analysis of Key Concepts

4.1 "Self"

In §3.1 I argued that Deep Self theorists need not be committed to the existence of a deep self in a way that commits them to the existence of some central, fundamental, all-important seat of agency or of some most authentic core of who a person is. In §3.2 I showed how conflating the relevant sense of the self for Deep Self theories of moral responsibility with some notion of agency *par excellence* only serves to intensify skepticism about alienation-based explanations of exemption.

Even if Deep Self theorists eschew these lofty commitments, though, the notion of 'self' needed for the theory still calls out for explanation. Fewer notions are as highly contested throughout the history of philosophy as the existence of some sort of self. However, I want to offer that Deep Self theorists can make use of a notion of the self that already rather uncontentiously exists while merely offering a reinterpretation of its boundaries.

All action is an interaction between an agent and her environment and/or circumstances. When I'm at the bottom of a stairwell, the action I end up performing is a function both of what I want to do and the fact that my options are shaped and constrained by the environment of the stairwell. While a first pass interpretation might draw the distinction between agent and environment at the bounds of the agent's body and the world around her, upon further reflection it seems that we sometimes are willing to think of even internal sensations as being part of the agent's environment rather than her agency. For example, if I have a very itchy elbow and someone offers me that they will donate $100 to famine relief if I refrain from scratching it, it seems that when I agentially navigate that situation the degree to which my elbow itches is part of the environment I navigate rather than part of my agency. The degree to which my elbow itches is part of the circumstantial factors that set the first-order normative facts of the situation—in this case, whether and to what degree it would be bad to scratch my elbow—rather than part of the *aretaic* evaluation of me as an agent.

Deep Self theorists can use the notion of 'self' they invoke to refer to the bounds between the agent and her environment. The suggestion that Deep Self theorists would then would be putting forth is that motivational states from which agents are wholly alienated are too much a mere function of the circumstances the agent finds herself in to count as being agential for the purposes of appraisal. Agents are not responsible for their actions if their actions are mere products of their environments. Notice that while we tend to be more in control of our selves than our environments, even physical features of our environments are often in our control to some extent. But when we are responsible for our environments it is only indirectly and only via our management of the parts of them that we have access to managing. Deep self theorists should want to say something like this about alienated motivational states as well.

4.2 "Internality"

With this analysis of self in place, I want to offer an analysis of the feature by which certain kinds of mental states are said to qualify as being part of the self. It is said by competing Deep Self theorists that all tokens of their favored mental state kind and no others bear the mark of internality. This is meant to mean something like the fact that any token instances of their favored mental state kind (whether endorsing, caring, valuing, or planning). So if, for example, valuing states bear the mark of internality, if you are motivated to return a lost wallet because you value it, this ensures that your motivation is internal. What, though, *is* this important property of 'internality' these deep self mental states are alleged to have that alien mental states lack such that they can 'speak for' the agent? In what sense can they speak for her, and what, exactly, do they say?

A surprisingly simple analysis of internality can help us make sense of the arguments made in favor of competing candidates for the relevant deep self mental states made by major Deep Self theorists in terms of it. My proposal is that a mental state kind bears the mark of internality iff it is the kind of state such that when an agent has a token mental state of that kind that disposes her to ϕ, there is sense in which the agent approves to some degree of her motivation to ϕ. So, for example, if it is true that caring states bear the mark of internality, this conditional would follow: if a person is motivated to call her mother because she cares about her, then she will approve to some degree of (i.e. like something about) her motivation to call her mother in this particular instance.

While I think it will be most helpful to stick with a largely intuitive understanding of what 'approving of' or 'liking' a motivation to some degree might mean, I do want to offer some clarifications. If an agent approves to some degree of her motivation to ϕ it does not mean that she merely prefers it over some horrible alternative motivation she might have otherwise had. She must take the prospect of being moved to action by that motivation to be appealing to some degree, even in the context of her *actual* and potentially conflicting motivations, even if its appeal is not, for her, decisive. This does not

necessarily mean that agent takes ϕing to be a good or well-justified course of action. Neither does it mean that the agent will be consciously aware of liking the motivation at the time of action. Our attitudes may not always be transparent to us, including our meta-attitudes, because we often simply do not take the time to reflect on our attitudes. For an agent to like one of her motivational states, she need only be such that she would take there to be something to acting on it in this circumstance if she were, at the time of action, to reflect on it (while holding fixed her other mental states).[31]

If I am right that this intuitive notion of approving to some extent of an action rather than merely being motivated to perform it is what makes the difference between cases in which we are willing to grant that an agent's action is caused by an process that bears the mark of internality and ones in which we are not, then we have located a common feature of any plausible candidate deep self mental state. Whether the deep self mental states are proposed to be endorsements, valuings, plans, or cares, or some disjunction of these, they succeed in guaranteeing agents' resultant actions will be internal by guaranteeing that the agent will approve of her action. This means we can locate a common analysis of internality that makes the debates among deep self theorists significantly less mysterious and metaphorical.

Elsewhere I have offered a Deep Self *account* of my own that offers both necessary and sufficient conditions for an agent's ϕing to be able to speak for her.[32] My account specifies the particular way in which an agent need approve of her action such that her approving guarantees that her action will be attributable to her in the relevant sense. But here my aim is different. It is merely to give a naturalistic *analysis* of the very concept of internality, understood as a shared necessary precondition for attributable agency by Deep Self theorists of different stripes.[33]

---

[31] While many Deep Self theorists write primarily in terms of whether or not a *mental state* is internal, whether the *resulting behavior* is internal or not is the question that is important for responsibility. For an action to be internal its causal mechanism needs to be suitably related to the fact that the person likes it—it cannot be wholly accidental. You might, for example, happen to really like the fact that you have a motivation that causes you to breathe but this does not make you responsible for breathing. (Thanks to a reviewer for the *Journal of Moral Philosophy* for helping me to clarify the scope and relevance of "liking" one's motivation.)

[32] Gorman (2019).

[33] While, as I have argued, lack of internality can do a good job of explaining certain kinds of urges and compulsions that can be genuinely troubling, Deep Self theorists ought to give up any potential aspirations of appealing to a lack of internality to provide an explanation for *all* desires that might be seen as pathological. One can persistently act in obsessive ways that hinder one's own interests while being motivated by desires that are internal in a thoroughgoing way. For example, while people with OCD tend to experience their compulsions as alien, making a lack of internality a good explanation for the fact that such behaviors are not attributable, people with OCPD sometimes fully embrace the targets of their obsessive motivations, believing, for example, that thoroughgoing oven-checking is appropriate and necessary for preventing fires. This analysis of 'internality' in terms of liking perhaps lays bare the futility of this more ambitious way one might hope to put the concept of 'internality' to use, but this project would

4.2.1 Endorsing Secures Agential Approval

It is perhaps easiest to see how approving of one's course of action will always be part of endorsing one's course of action on Harry Frankfurt's endorsement view. Second-order volitions are meant to secure the fact that the agent is not only motivated to act in the way that she does but that she is personally invested in that particular course of action. This aspect of the endorsement view comes out particularly clearly in Frankfurt's discussion of the contrast between wontons and full-fledged agents who form second-order volitions.

For Frankfurt, our actions are attributable to use because we are not wantons, people who let their strongest motivational states move us to action irrespective of any opinion we might have on the matter.[34] Why, on Frankfurt's view, are we meant to think that the wanton's actions aren't attributable to her in the relevant sense? For the wanton,

> …it makes no difference to him whether his craving or his aversion gets the upper hand. He has no stake in the conflict between them and so…he can neither win nor lose the struggle in which he is engaged.[35]

This means that when an agent *is* responsible for her action it is at least partially because she "has a stake" in the outcome of the conflict among the economy of her desires. Having a stake in the conflict between first-order desires competing to become an effective desire seems to amount to having an opinion on the outcome. In other words, the agent needs to approve of being motivated to act in the way that she does in order for her action to bear the mark of internality.

4.2.2 Valuing Secures Agential Approval

Approving of one's action is also key to the valuing version of the deep self view, although a mistaken picture of the contrast between valuing and desiring at work in the theory may make this idea seem somewhat obscure. There is a picture of human agency that pits what an agent wants to do against what she thinks would be best to do, conceiving of the two things as wholly separate. On this view it is nice when an agent is motivated to do what she thinks is best, but this is either accidental or caused by the agent bringing her motivations in line with what is best; it is not that there is any motivational force to her judgment that a certain course of action is best. Given this sort of picture, it

---

always be destined to fail. It is just a fact that people can, unfortunately, not just like but also genuinely care about, plan, intend, and value the content of desires that are self-undermining. If we want to exempt such people from responsibility, we will have to make reference to a further fact, such as their epistemic status, not just their agential status. For further discussion see [Gorman, Forthcoming].

[34] Frankfurt (1988): 19.

[35] Frankfurt (1988): 89.

would be hard to see how the fact that an agent values some course of action would be sufficient to guarantee that she personally approves of it in the right kind of sense to make it self-expressive. Valuing, however, is often held to have some more intimate connection with motivation. And once this is granted, it is easier to see the connection with approval.

To act on one's valuing state in the sense that defenders of the valuing view conceive of it is never to merely act in accordance with what one coincidentally believes to be good. Rather, valuing is thought to have something to do with agency by issuing from a faculty that has a particular sort of "grip" on her motivations. If the thought "it's the right thing to do" is meant to have a grip on motivation, it must be because the second thought, "and I approve of doing the right thing," is also present in some form. Whether the second thought is a matter of the meaning of rightness, a truth about human nature, or a standing disposition that happens to be present in agents like us (or something else), the fact that the agent approves of acting as she does because it is right seems baked into the story. Watson explains that the sort of motivational power exerted by valuing is special because we are concerned to bring about the satisfaction of desired ends for some reason that goes beyond the fact that acting alleviates the suffering of having the unsatisfied desire. For an agent to value φ-ing is for her not just to desire to φ but to set φ-ing as an end for herself. And so an agent must not only be motivated to φ, but be motivated in the special way that comes about from approving of the end to which φ-ing aims such that it gives you a reason to φ.[36] And so, on the valuing view, valuing is the relevant deep self mental state precisely because it guarantees that the agent's effective desire becomes effective because she approves of her course of action.

4.2.3 Planning Secures Agential Approval

According to the planning view, an agent is responsible iff she acts in accordance with her policy about how to act in such a situation.[37] Her policy-setting may be governed by her values in many cases, and in those cases the same considerations I raised regarding the valuing view apply. But in cases that are normatively underdetermined, she forms or acts on a previously determined policy that she just *decides* to treat as reason-giving. If agents in these cases just follow personal policies that are not governed by anything as strong as all-things-considered judgments about what would be best, it might be far from clear that agents who act in accordance with these policies need approve of their actions.

---

[36] Watson (1975): 210-211.

[37] In the article I draw from here, Bratman specifically brackets off questions of responsibility, focusing his discussion on identification alone: "…I want to see if we can, instead, describe without independent appeal to judgments of responsibility—a fairly unified phenomenon that is plausibly seen as the target of such talk of identification" (Bratman [1996]: 2) His picture might just as easily be considered as a contending deep self account of attributional-responsibility, however, and, with this caveat, I will proceed as though it were put forth as one.

However, following Velleman, Bratman acknowledges the possibility of a case in which an agent forms a plan in such a detached way that the action she takes when she fails to act in accordance with it would still be attributable to her. This provides a reason to supplement the story about what must obtain in these sorts of cases for the agent to be responsible. Bratman supplements his account by adding that the agent who φs must be *satisfied* with her decision to treat her desire to φ as reason-giving not only when she decides, but also "when the chips are down" at the time of action.[38] If an agent meets this condition, it seems to me that she would have to approve of at least something about it.

Bratman understands satisfaction with a policy not in terms of the presence of a particular attitude, but rather, in terms of the alignment and integration of the policy with the agent's other policies: "One is satisfied with such a decision when one's will is, in the relevant ways, not divided: the decision to treat as reason-giving does not conflict with other standing decisions and policies about which desires to treat as reason-giving."[39] But notice that this analysis of satisfaction only makes sense as an analysis of satisfaction when we think of the sum of the agent's other policies as providing a guide to what the agent generally approves of doing. Again, here, agential approval of some form seems to drive intuitions about whether or not the state is identified with in such a way that it truly bears the mark of internality.

4.2.4 Caring Secures Agential Approval

Sripada's caring view identifies *cares*, a *sui generis* mental state with a particular profile of emotional, judgmental, motivational, and commitmental dispositional tendencies as the relevant kind of deep self mental state. While this makes giving an analysis of what makes all and only those mental states count as internal according to the caring Deep Self theorist, I do think that the sense of approval of one's action that I have identified is consistent with being a necessary condition on acting in the promotion of one's care, given the way Sripada characterizes cares.

Notably, "approval" is explicitly listed as one of the emotions brought about by acting in accordance with ones cares.[40] While cares also involve valuing, Sripada argues *contra* the valuing view that only some subset of an agent's actions that are motivated by evaluative judgments bear the mark of internality—those that bear the right dispositional tie to one's cares. These valuings are properly internal because they cast some end to which the action aims in a favorable light for the agent; it is only when doing what one takes to be the best justified course of action happens to *matter* to the agent on some personal level that it is properly internal. This sense of mattering seems to be fundamentally tied to approving. Finally, the motivational and commitmental aspects of caring seem to implicate at least some degree of approval. For Sripada, if an agent cares about X, she is

[38] Bratman (1999): 202.

[39] Bratman (1996): 201.

[40] Sripada (2016): 8.

intrinsically motivated to perform actions that promote the achievement of X and will want to continue caring about X. Together, this makes it the case that when φ-ing promotes the achievement of X, and X is something the agent cares about, the agent has an intrinsic desire with a positively valenced higher-order attitude towards being motivated by it. Although the consideration of the promotion of the achievement of X may not outweigh other factors in a given case, the agent still would seem to have to approve of being motivated to φ in at least a minimal or *pro tanto* sense for this to be the case.

Each of the major candidate deep self mental states, it seems, bears the mark of internality *by* making it the case that the agent will necessarily approve of or like her action to some degree. And so it seems Deep Self theorists can adopt a surprisingly simple analysis for the concept of internality, one which can simplify their message as well as clarify debates between different Deep Self theorists.

5 Conclusion

To sum up, I have shown that there are quite naturalistic motivations to adopt a Deep Self view, I have diffused some of the main sources of initial skepticism, and have given non-mysterious analyses of the concepts of "self" and "internality" with the aim of bringing Deep Self views out of the metaphorical fog and down to earth. In addition to highlighting the naturalistic motivations and providing naturalistic analyses of Deep Self concepts, I have suggested that Deep Self theorists who wish to persuade skeptics that the family of views does more than offer wrongdoers a fortress of moral evasion should (a) be careful to clarify their commitments regarding what they mean by talk of a "deep" self, and (b) make a top priority of differentiating compulsion from weakness of will. In turn, I argue that critics should not preemptively dismiss Deep Self views based on common mischaracterizations of their commitments. In addition, insofar as arguments for Deep Self views rely on the veridicality of testimony about experiences of alienation, this testimony should be taken seriously. While the veridicality of such experiences is surely open to debate, the wholesale dismissal of the phenomenon should be carefully scrutinized, as it may be influenced by the confluence of testimonial and hermeneutic injustice.[41]

In this article I have offered more of an *apologia* than a decisive argument in favor of Deep Self views, as there are certainly plenty of criticisms I have left unaddressed. But

[41] One way to characterize this might be in terms of José Medina's concept of epistemic responsibility. Critics should maintain humility with respect to their epistemic limits regarding the issue and a stance of curiosity/diligence in closing the epistemic gaps (Medina [2013, pg. 42-3]).

what I do hope to have established is that you can be a Deep Self theorist while simultaneously valuing clarity and naturalism, and that it would be an ironic mistake to summarily dismiss this family of views as metaphysically incoherent or as morally suspect.

Bibliography

Arpaly, Nomy (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. Oxford University Press.

Arpaly, Nomy & Schroeder, Timothy (1999). Praise, Blame and the Whole Self. *Philosophical Studies* 93 (2):161-188.

Björnsson, Gunnar & Pereboom, Derk (2016). Traditional and Experimental Approaches to Free Will and Moral Responsibility. In Justin Sytsma & Wesley Buckwalter (eds.), *Companion to Experimental Philosophy*. Blackwell.

Bratman, Michael (1996). Identification, Decision, and Treating as a Reason. *Philosophical Topics* 24 (2):1-18.

Bratman, Michael (1999). *Faces of Intention: Selected Essays on Intention and Agency*. Cambridge University Press.

Bratman, Michael (2003). A desire of one's own. *Journal of Philosophy* 100 (5):221-42.

Braut, Jennifer (2018). "Investigating Misophonia: A Review of the Empirical Literature, Clinical Implications, and a Research Agenda." *Frontiers in Neuroscience.* 12

Buss, Sarah (2012). Autonomous Action: Self-Determination in the Passive Mode. *Ethics* 122 (4):647-691.

Calhoun, Cheshire (2008). "Losing One's Self." In *Practical Identity and Narrative Agency*, ed. Catriona Mackenzie and Kim Atkins. London: Routledge.

Carel, Havi, and Kidd, Ian James. 2014. "Epistemic Injustice in Healthcare: A Philosophical Analysis." *Medical Health Care and Philosophy* 17: 529–40.

Fischer, John Martin (2010). The Frankfurt cases: The moral of the stories. *Philosophical Review* 119 (3):315-336.

Flynn, Abi (2017). Misophonia and Me – A massive pouring out of my heart and soul. It's long. Blog post. *https://abiflynncancerblog.wordpress.com/2017/02/03/misophonia-and-me-a-massive-pouring-out-of-my-heart-and-soul-its-long/*

Frankfurt, Harry (1987). Identification and Wholeheartedness. In Ferdinand David Schoeman (ed.), Responsiblity, Character, and the Emotions: New Essays in Moral Psychology. Cambridge University Press

Frankfurt, Harry (1992). The Faintest Passion. *Proceedings and Addresses of the American Philosophical Association* 66 (3):5-16.

Frankfurt, Harry (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66 (23):829.

Frankfurt, Harry (1971). Freedom of the will and the concept of a person. *Journal of Philosophy* 68 (1):5-20.

Frankfurt, Harry (1988). *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.

Frankfurt, Harry (2006). *Taking Ourselves Seriously & Getting It Right*. Stanford University Press.

Fricker, Miranda (2007). *Epistemic Injustice: Power and the Ethics of Knowing*, Oxford: Oxford University Press, 17–29.

Germiniani, Francisco et. al (2012). Tourette's syndrome: from demonic possession and psychoanalysis to the discovery of gene. *Arquivos de Neuro-Psiquiatria*, *70*(7), 547-549.

Gorman, August (2019). The Minimal Approval View of Attributability. In David Shoemaker (ed.), *Oxford Studies in Agency and Responsibility 6*. Oxford University Press.

Gorman, August (forthcoming). What is the Difference Between Weakness of Will and Compulsion? *Journal of the American Philosophical Association*.

Gosselin, Abigail (2019). Philosophizing from Experience: First-Person Accounts and Epistemic Justice. *Journal of Social Philosophy* 50 (1):45-68.

Hanagarne, John (2014). *The World's Strongest Librarian: A Book Lover's Adventures.* NYC: Gotham Books.

Kumar, Sukhbinder et al. "The Brain Basis for Misophonia." *Current Biology*27.4 (2017): 527–533. *PMC*. Web. 14 Sept. 2017.

Jackson, Jake (2017). "Patronizing Depression: Epistemic Injustice, Stigmatizing Attitudes, and the Need for Empathy." *Journal of Social Philosophy* 48: 359-376.

Jackson, L, et al. An exploration of the social identity of mental health inpatient service users. *Journal of Psychiatric and Mental Health Nursing*. 2009;16(2):167-176.

Kurs, Rena and Grinshpoon, Alexander (2018). Vulnerability of Individuals With Mental Disorders to Epistemic Injustice in Both Clinical and Social Domains, *Ethics & Behavior* 28:4, 336-346.

Levy, Neil (2011). Expressing who we are: Moral responsibility and awareness of our reasons for action. *Analytic Philosophy* 52 (4):243-261.

Lippert-Rasmussen, Kasper (2003). Identification and responsibility. *Ethical Theory and Moral Practice* 6 (4):349-376.

Matheson, Benjamin (2018). The Threat from Manipulation Arguments. *American Philosophical Quarterly* 55 (1): 37-50.

McKenna, Michael (2008). Frankfurt's argument against alternative possibilities: Looking beyond the examples. *Noûs* 42 (4): 770-793.

Medina, José (2013). *The Epistemology of Resistance: Gender and Racial Oppression, Epistemic Injustice, and Resistant Imaginations.* Oxford: Oxford University Press.

Mele, Alfred (2006). *Free Will and Luck*. Oxford University Press.

Mitchell-Yellin, Benjamin (2014). In Defense of the Platonic Model: A Reply to Buss. *Ethics* 124 (2):342-357.

Mitchell-Yellin, Benjamin (2015). The Platonic model: statement, clarification and defense. *Philosophical Explorations* 18 (3):378-392.

Moran, Richard. (Forthcoming). Christine Korsgaard Festschrift.

Penelhum, Terence (1971). The importance of self-identity. *Journal of Philosophy* 68 (October):667-78.

Sartorio, Carolina (2016). A Partial Defense of the Actual-Sequence Model of Freedom. *The Journal of Ethics* 20 (1-3):107-120.

Scheler, Max. 2008. The Nature of Sympathy, translated by Peter Heath. New Brunswick, NJ: Transaction Publishers.

Schroeder, Timothy & Arpaly, Nomy (1999). Alienation and externality. *Canadian Journal of Philosophy* 29 (3):371-387.

Shoemaker, David. (2015a). "Ecumenical Attributability" in *The Nature of Moral Responsibility* ed. Randolph Clarke, Michael McKenna, and Angela Smith. (Oxford: Oxford University Press.)

Shoemaker, David (2015b). *Responsibility From the Margins*. Oxford University Press.

Sripada, Chandra. "At the Center of Agency, the Deep Self" Unpublished Manuscript. *https://umich.app.box.com/s/vlknn6s4ggnc7tuefft2*

Sripada, Chandra (2016). Self-expression: a deep self theory of moral responsibility. *Philosophical Studies* 173 (5):1203-1232.

Sripada, Chandra (2017). Frankfurt's Unwilling and Willing Addicts. *Mind* 126 (503):781-815.

Vargas, Manuel (2011). Revisionist Accounts of Free Will: Origins, Varieties, and Challenges. In Robert Kane (ed.), *Oxford Handbook on Free Will, 2nd Edition*. Oxford University Press.

Velleman, David (1992). What Happens When Someone Acts? *Mind* 101 (403):461-481.

Watson, Gary (1975). Free agency. *Journal of Philosophy* 72 (April):205-20.

Wolf, Susan (1987). Sanity and the Metaphysics of Responsibility. In Ferdinand David Schoeman (ed.), *Responsibility, Character, and the Emotions: New Essays in Moral Psychology*. Cambridge University Press. pp. 46-62.

Wolf, Susan (1990). *Freedom Within Reason*. Oxford University Press.