

The Minimal Approval Account of Attributability¹

August Gorman

Abstract: This paper advances a new agentially undemanding account of the conditions of attributability, the Minimal Approval account, and argues that it has a number of advantages over traditional Deep Self theories, including the way in which it handles agents with conditions like addiction, Tourette syndrome, and misophonia. It is argued that in order for an agent to be attributionally-responsible, the mental process that leads to her action must dispose her to be such that she would, upon reflection, approve to some minimal degree of being moved to action by the motive on which she in fact acts.

Keywords: Attributability, Moral Responsibility, Deep Self, Alienation, Tourette syndrome, Misophonia

I. INTRODUCTION

In “Free Agency,” Gary Watson writes, “normally, in the pursuit of the objects of our wants we are not attempting chiefly to relieve ourselves. We aim to satisfy, not just eliminate, desire” (Watson 1975: 20). This is a crucial insight into the way in which we are related to our actions, the way in which they are the kinds of things that can be (or can fail to be) attributable to us. While Watson takes himself to be describing just one aspect of a more complex state that he thinks is required for attributability, I take this idea to be the defining feature of attributability. In this paper I take this intuitive idea, that one’s action must not merely be caused by a rogue motivation or a desire to eliminate some rogue motivation, and flesh it out into a complete theory of the conditions for attributability.

The project of finding the proper conditions on which an agent’s action can be attributable to her is the project of delimiting a distinct category of acts that properly stem from a person’s own agency. The goal is to find a filter that separates the sorts of acts that an agent cannot truthfully claim come from ‘outside themselves’ from the ones that stem from mere neurological noise, brain activity that circumvents the processes we take to involve the self. This task is important because this class of acts is the class of acts for which agents are at least in principle morally responsible.

¹ I owe thanks to David Shoemaker, Gary Watson, Steve Finlay, Eugene Chislenko, Renee Jorgensen Bollinger, and Nathan Robert Howard for comments on earlier drafts, and to Philip Swenson, Neil Tognazzini, Justin Coates, Gunnar Björnsson, Aness Webster, Matt Talbert, Eric Wiland, Elizabeth Harman, Santiago Amaya, Andreas Brekke Carlsson, Anneli Jefferson, Noel Dominguez, Facundo Alonzo, Ben Mitchell-Yellin, Janet Levin, Ralph Wedgwood, Chuck Goldhaber and Sam Cirulis, for helpful feedback on earlier formulations of the ideas in this paper. The paper also greatly benefitted from discussion by the participants of the fourth New Orleans Workshop on Agency and Responsibility, Temple University’s Freedom and Responsibility Fall 2017 Seminar, and USC’s Spring 2016 Dissertation Seminar. I am especially grateful for Jonathan Quong’s detailed comments on drafts and endless hours of discussion throughout my development of these ideas.

In setting out my theory here, I remain agnostic about the proper route from attributability to blameworthiness, and simply understand attributability as a necessary condition for the familiar kind of blameworthiness that moves from an agent’s wrongdoing to the fact that some sort of response *to the agent* is merited. This leaves open the possibility that there are additional conditions that agents must meet in order to be responsible, such as an epistemic or normative competence condition, etc. When an agent stands in the proper agential relationship to her action such that it opens her up in principle to being morally responsible for the act (or what it reveals about her), I will describe such an agent as being “attributionally-responsible” for her action.

I motivate my project by first briefly motivating the appeal of Deep Self views, and then showing how while these views each give a plausible sufficient condition for attributability, they fail to locate a necessary condition. I then present my new account of attributability that aims to solve this problem: the Minimal Approval Account. I motivate three distinctive components of the view: it requires only partial rather than wholehearted endorsement of action, it requires merely hypothetical rather than actual endorsement, and it requires a special kind of endorsement: the agent must endorse her action because she approves of its corresponding desire’s satisfaction. I then turn to two advantages of the Minimal Approval Account. First, it is able to explain that human agents, unlike nonhuman animals, are candidates for attributional-responsibility because of special agential capacities, while allowing that humans regularly act without first engaging in explicit reflective processes. Second, it does not tie attributability too closely to any particular causal story about action-production. I conclude with a discussion of how the theory fits in to the broader landscape of the literature.

II. MOTIVATION FOR THE PROJECT

A. *The Appeal of Deep Self Views*

In contrast to control-based views of the conditions under which an agent is responsible for her action, Deep Self views focus not on an agent’s ability to do otherwise, but rather, on the combination of mental states that lead to the agent’s action. They hold that an agent’s effective motivation (motivational state that in fact leads her to action) must have the proper mesh with her other psychological states for her action to be able to “speak for her” by expressing what she stands in favor of doing.² When an agent is not attributionally-responsible for her behavior, according to Deep Self theorists, the reason is because she fails to agentially identify, or stand behind, her motivation in a way that alienates her from its results. When an agent, instead, stands in favor of the actions she undertakes, it seems reasonable to hold that her actions express something about her

² One source of potential confusion here is that views that posit that *control* over self-disclosing agential capacities as a condition on moral responsibility are sometimes referred to as Deep Self views. Agnieszka Jaworska refers to these views as “Broad Identificationist Views” in contrast to “Narrow Identificationist Views” on which an agent must actually be motivated by the relevant mental states (see Jaworska 2015). My own tendency is to think of control-based views as species of Classical Compatibilism that posit that the relevant sense of alternative possibilities to ground moral responsibility has to do with dispositions to act differently if one had chosen/endorsed/valued differently and to not include them under the heading of “Deep Self views.”

agential character. And taking agents' actions to reveal their character, many think, plays some crucial role in our practices of holding each other moral responsible.³

Deep Self views hold that only some subset of an agent's motivational states can produce actions for which she is attributionally-responsible, but differ in which subset they take to be relevant. On most extant views of this type, in order for an agent's action to speak for her, the motivational state that she acts on must be or be caused by a special mental state. The most popular candidates are second-order volitions, valuings, plans or commitments, and caring states.⁴ So, for example, on Frankfurt's second-order volition view, in order for an action to be attributable, the agent must decide which of her first-order desires she wants to act on and actually act in accordance with this. On Watson's valuing view, an agent who satisfies the conditions for attributability is motivated by a valuing state to perform the action that she takes to be all-things-considered best. These particular kinds of mental state types are often singled out because they are thought by their proponents to be made special by the fact that they are *invariably internal*, that is, no agent could be alienated from states of these types.

Deep Self views represent one of the most plausible alternatives to control-based views of the conditions of moral responsibility. They forge past a stalemate in the literature regarding how best to understand the "ability to do otherwise," by insisting that the real criterion for moral responsibility lies in the factors that actually help to explain why an agent in fact acts in the way that she does.⁵ In addition, Deep Self views help to explain a real aspect of the phenomenology of our agential lives that calls out for explanation: humans sometimes fail to identify in the usual way with the causal springs of their actions in a way that seems to undermine agency.

While this phenomenon may not be familiar to all, it often plays a large role in the lives of people suffering from compulsions, addictions, and phobias. In our actual practices we do often exempt people suffering from such conditions from (at least certain kinds of) moral responsibility and people who suffer from these conditions in real life often feel a certain kind of alienation from the source of their effective motivations. Deep Self theorists posit that this is no coincidence, since these feelings are often veridical. As Frankfurt writes, declarations by agents that their actions are external "may be, of course, shabbily insincere devices for obtaining unmerited indulgence. Or they may be nothing more than emphatic expressions of regret. But it is also possible that they are genuinely descriptive" (Frankfurt 82: 63). Deep Self theorists take this possibility seriously, and then look for the conditions under which it would make sense to say that an agent's action really *doesn't* speak for her.

B. Failure to Locate a Necessary Condition on Attributability

³ For example, some think the deep self is important for picking out the domain of the 'will' in quality of will accounts, some think it enables aretaic evaluation of an agent which is itself a form of responsibility, and others still take revealing one's character to be a precondition for being an apt target of the reactive attitudes.

⁴ Some of the key works in which such views are put forward include Frankfurt (1971), Watson (1979), Bratman (1996), Shoemaker (2003), and Sripada (2015). Shoemaker (2015) puts forth a view on which an attributionally-responsible agent's motivational state may mesh *either* with her cares or commitments.

⁵ See Frankfurt (1969), Frankfurt (1988), Mele (2006), McKenna (2008), and Sartorio (2016) for discussion.

Despite the advantages of Deep Self views, a well-rehearsed objection is that while these views tend to locate promising sufficient conditions for attributability, they fail to locate necessary conditions.⁶ For example, traditional Deep Self views seem unable to respect the intuitive judgment that persons can be responsible when suffering from weakness of will.⁷

Suppose there is a murderer who is conflicted between going out on a killing spree, a pastime he loves and treats like a hobby, and getting his morning gym workout in so he doesn't fall off of his routine. After considering and giving some weight to each option, he decides that acting on his desire to work out is what he most wants to do, it aligns best with his values, and is consistent with the plans he has set for himself. However, in the end, he is overcome by his wish to continue honing his murderous craft. His killing spree would certainly seem to reveal something about what he is like such that his action would speak for him for the purposes of attributional-responsibility; he is clearly blameworthy despite the fact that he endorses, in the senses relevant for most Deep Self accounts, going to the gym at the time of action.⁸ To illustrate the problem, on Frankfurt's view he forms a second-order volition to go to the gym but some other first-order desire becomes his effective motivation. But if Frankfurt's account were the right account of attributional-responsibility, we would have to say that he is not attributionally-responsible for his killing spree. But this seems clearly to be the wrong result.

This is a particularly difficult problem for traditional Deep Self views because they are specifically designed to show how acting on desires that agents themselves do not see as most favorable undermine agency in such a way as to make such actions non-attributable, as in the case of compulsion. But weakness of will is usually described as the failure to act in accordance with what one finds to be the most favorable course of action, and yet intuitively we think weak-willed actions are attributable. In short, they lack the resources to differentiate compulsion from weakness and in exempting compulsive action they overextend to exempt weak-willed action. Since weak-willed actions do seem attributable, but traditional Deep Self psychological state requirements for attributable action are not compatible with this, these views fail to accurately identify the necessary conditions for attributability.

III. THE ACCOUNT

A. *Partial Endorsement*

⁶ This objection can be motivated in several different ways. Here I focus on weakness of will cases, but similar arguments can be run by focusing on spontaneous or unreflective action.

⁷ This objection has been raised, in various forms, by Vihvelin (1994), Haji (1998), Haji (2002), Fischer (2012a), Fischer (2012b), McKenna (2011), McKenna and van Schoelandt (2015), and Strabbing (2016). A complication that should be noted here is that not all Deep Self views are presented as views of the conditions for responsibility, and some theorists propose these sorts of conditions merely as views of self-governance, intentional action, autonomy, or agency *par excellence*. This criticism should be taken to apply only to the (many) Deep Self theorists who do take acting in accordance with one's Deep Self to be a criterion for responsibility.

⁸ In my paper "How Should Deep Self Theorists Account for Weakness of Will?" (Gorman, Unpublished) I address other possible responses to the weakness of will problem for Deep Self theorists. While I am skeptical that any other solutions are viable, in *this* paper I only want to make the point that the problem of weakness of will, on the face of it, gives us at least some reason to look for a new or significantly modified account.

The aforementioned considerations give us some reason to question the viability of traditional Deep Self accounts. In order to preserve the general idea of a Deep Self account in the face of this kind of criticism, we might need to radically rethink the kind of thing that might be required for an agent's action to count as attributable. One route forward is this: while it makes sense to say that in compulsive cases, agents are wholly alienated from their actions, it seems that there are cases, like the case of the weak-willed murderer, in which an agent doesn't stand behind his action as being *the* thing to do and yet still endorses it, in at least a partial way. While most extant Deep Self views give accounts of what would be required for an agent to *fully* stand behind her action, the lesson I think Deep Self theorists should take from thinking about cases of weakness of will is that an agent does not need to wholeheartedly endorse her course of action in order to be responsible for it. If lack of identification with one's action is to be a relevant consideration in exempting an agent from attributional-responsibility, we'll need to understand this in terms of total lack of identification with one's action, rather than less-than-complete identification with one's action.⁹

We can then ask the following question: what are the minimal conditions for an agent's action being attributable, such that failing to meet these conditions would completely alienate her from her action? If, as I'm proposing, an agent need only identify with her course of action *in some way* or *partially* for her action to reveal something about her such that it speaks for her in the sense relevant to attributional-responsibility, we'll need a way to understand partial identification. In order to make progress here it will be helpful to borrow some terminology from Harry Frankfurt. In Frankfurt's terms, a second-order desire is a desire to desire something, and a second-order volition is a desire that one of your first-order desires be motivationally effective, that is, actually propel you into action (Frankfurt 1971: 15-16.) To fully endorse an action, an agent forms a second-order volition to act on a desire to φ and actually does φ . But in deliberating about which first-order desire to act on, we are often conflicted.

In order to see the difference between what I'll call complete and merely partial endorsement, it will be easiest first to look at a case in which an agent feels herself being pulled in more than one direction by her first-order desires and then explicitly deliberates about what to do. Consider the following case:

THREE DESIRE THERESA: Theresa is currently at work and has three first-order desires, each with the potential to pull her in a different direction: she wants to complete the assignment she has been given by her boss, she wants to spend her time writing some thank you emails to some relatives who are waiting to hear from her, and she wants to slap her boss. She then considers each of these desires and upon consideration she gives some weight to acting on

⁹ This is not to deny the possibility that degree of identification with one's action may play some further role in determining features of blameworthiness. It seems at least *prima facie* intuitive that wholehearted embrace of morally wrong action could be cause for extra/more intense blame or blame of a special kind of character, and my proposal should not be taken as being incompatible with this. I am merely suggesting that we shift from a conception in which anything less than wholehearted identification falls short of attributability to one in which merely partial identification makes the cut, so to speak.

her desire to get her work assignment done, and some weight to acting on her desire to send her thank you emails. Slapping her boss, she realizes, isn't even a contender; even though she is struck by the raw urge to do it, she would never *really* want to do this. She's not even angry with her; it's just an occasional urge that pops into her mind. She decides, in the end, to work on her assignment.

It's important to note that the deliberation process I'm imagining is distinct from *rational* deliberation about what action would be *justified*. While this is controversial, it seems to me that our actions often clearly meet attributability conditions even when we do things that we take to be wholly unjustifiable. For example, I can easily imagine a case in which a person commits a crime, which she knows to be morally wrong, in order to help an unworthy ex-lover, out of residual love for him. It seems clear that her action is attributable to her even though it seems possible that she might see her action as wholly unjustifiable.¹⁰

Three possible outcomes in this scenario lead to three different levels of potential alienation. First, there is the outcome in which what she actually does is her work assignment. She feels no sense of alienation in this case, given that she acts on the desire of hers that she most wants to act on; she stands behind her action.

Next, there is the outcome in which, when she opens up her work assignment, she shifts over to her email and instead starts writing those thank you notes. It's possible this may feel somewhat alienating to her, as she is not, in the end, motivated by the desire that she decided she most wanted to act on. However, acting on a desire to write those thank you emails is in line with at least something that she wanted for herself. She may have decided in the end that what she wanted for herself *more* was to act on her desire to work on her assignment, but nevertheless she did want to some degree to act on her desire to write the thank-you emails. Furthermore, let's make the reasonable stipulation that it was no mere coincidence that she gave some weight to her desire to write thank you notes and the fact that she actually wrote thank you notes. The elements of her psychology that led her to give some post-reflective weight to her desire to act on her desire to write thank you notes were also involved in causing her to actually write the thank you notes. This act, I want to argue, is therefore not wholly alien to her since she at least partially endorses it.

Finally, there is the outcome in which she, despite never seriously considering it as a contender for the motivation she would really want to act on, slaps her boss. If this happens, something seems to have gone seriously awry. Even though she had the first-order motivation to slap her boss, she experiences complete alienation since she acts on a motivation wholly outside of what was even in contention for what she wanted for herself to do upon reflection. In this case

¹⁰ See Shoemaker (2011: 610-611) for further discussion of this point. Silverstein (2017) also gives a compelling case that practical reasoning, understood as reasoning about what to do and normative reasoning, understood as reasoning about what one ought to do, are non-identical. An agent can settle the question of whether or not she should dine out for lunch while deciding to forestall reasoning about whether or not she will in fact do so until tomorrow. In this case normative reasoning has finished long before practical reasoning has even begun. While this by itself does not foreclose the possibility that attributable agents must act in accordance with reason, it does drive a further wedge into the supposed tight connection between valuing and ordinary attributable action.

she is moved to action by a bit of her psychology that stands entirely outside of the complex psychological dispositions that affect choices about which motivations she would want herself to act on. Being moved to action when one would fail to even partially endorse its motive is, I contend, what makes actions that are compulsive in this way stand outside of one's agency, and thus fail to be able to speak for the agent. So, in order for an action to be able to speak for an agent it must be suitably related to the motivations the agent would at least give some second-order weight to in deciding on which motivation to act.¹¹ Returning briefly to the case of the weak-willed murderer, the notion of partial endorsement now gives us the resources to say that he *is* attributionally-responsible for his killing spree, as he gives some weight to his desire to go on such a killing spree.

B. *Hypothetical Versus Explicit Endorsement*

Other endorsement-based views posit that it is the actual *act* of endorsement that makes the agent's resultant action belong to her; the act of taking a stand results in ownership of one's motivation and its results.¹² However, not every case of attributable action involves the amount of intra-psychological reflection of Theresa or the weak-willed murderer. In fact, we rarely go through the explicit process of taking a stand on which of our desires to act on in the course of deciding how to act. Yet, we are still able to say with some degree of confidence whether or not our actions align with the motivations we would have wanted for ourselves to act on if we *did* consider which motivations to act on. And this is what seems to make the difference between whether or not we identify at all with our actions. It is the fact that what the agent to some degree would endorse and what she in fact does align with one another, I propose, rather than any actual act of endorsement, that makes it the case that an action is suitably related to an agent's deep self such that it can speak for her.¹³ Since what's important here are the psychological dispositions and not the endorsement itself, the actual act of endorsement can be merely hypothetical. So the view, so far, is this. An agent's action can speak for her iff her act is suitably related to the fact that she would at least partially endorse (give some weight to) the desire that she acts on, if she were to consider which of her motivations she wanted to act on.

We frequently do not explicitly deliberate about or give additional weight to our competing first-order motivations *en route* to acting, and so views that require explicit endorsement predict that we will regularly fail to identify with the springs of our actions. However, our actions are

¹¹ I take it that Bratman's notion of "being satisfied with a decision to treat the desire as reason-giving" may be sufficient for this kind of second-order weight-giving, but it is not necessary (Bratman 1999: 202). It's plausible that an agent must give some post-reflective weight to acting on her desire in order to be satisfied with a decision to treat it as reason-giving. But Bratman's criterion is not necessary for the kind of weight-giving I'm interested in here, since the agent need not take the desire to be reason-giving, and rather may just desire to act on it to some degree when she considers it as a contender for effective desire in the given situation (even if she does not see it in any way as providing her with a justificatory *reason* in any sense to act on the desire that she does).

¹² This is, for example, crucial to David Velleman's understanding of Frankfurt's view. See Velleman (1992).

¹³ John Doris considers a similar proposal, writing that "The appropriate standard is counterfactual: Identification may be said to obtain if a person *would have* identified with the determinative motive of her behavior at the time of performance had she subjected it to reflective scrutiny. Accordingly, unreflective persons – as all of us are sometimes – may be quite legitimately responsible for their unreflective behaviors." Doris (2002), 141.

almost always caused by the kinds of motivations to which we would give at least *some* weight if we were to deliberate, and in the rare cases in which a person's action fails to be produced by such a motivation, we generally attribute the result to some sort of dysfunction. In moving from explicit endorsement to hypothetical endorsement, the view entails that relatively few actions fail to meet the requirements of attributability.

I think this constitutes a point in favor of the hypothetical view because it can explain why some objectors to traditional Deep Self views are tempted to say that people who claim to be alienated from their actions must be making excuses. Objectors to traditional Deep Self views sometimes object that they can too easily imagine situations in which they themselves might not go through the proposed attributability-granting processes and even, perhaps in hindsight, feel *some* sense of alienation from their actions, while they nevertheless maintain strong intuitions that they are attributionally-responsible for those actions. But on a hypothetical-partial-endorsement-based view, complete alienation of the kind that undermines attributability is rare enough that it is entirely possible that these objectors seldom meet the conditions for it.¹⁴

C. Approval: Endorsement with a Further Aim than Elimination

On the view so far under consideration, Frankfurt's unwilling addict, understood as someone who would give no post-reflective weight to his desire to take drugs at the time of action, is another perfect example of a person whose action does not speak for him. Such a view also gives the verdict that willing addicts, even those who are merely partially willing at the time of action, are responsible for their taking drugs. That is, an agent who comes to take a drug due to a chain of mental states that ensure that she, at the time of action, would give some post-reflective weight to her desire to take drugs, acts in such a way that is self-disclosing for the purposes of attributional-responsibility.¹⁵ When we fill in the story in a certain kind of way, this seems to be

¹⁴ Despite these considerable benefits, some may protest that hypothetical endorsement is a poor substitute for actual endorsement. In an analogous discussion about hypothetical consent in law and applied ethics, David Enoch makes the point that in the case of consent, actual consent is sometimes important for reasons of sovereignty and sometimes for reasons of non-alienation. Hypothetical consent can answer to the latter but not the former concern. Something very similar seems to me to be true of endorsement. There are certain ways of thinking about actual endorsement that cast it as granting motivations with special agential power. For example, if the aim is to "put the agent back into the picture" of the causal story of action, then it's clear hypothetical endorsement won't do. But, later in this paper I will offer some cases in which non-attributable agents seem to make choices for themselves about what to do via processes that are very similar to the processes of attributable agents. This casts doubt on the idea that what goes awry in non-attributable actions is that the agent is missing from the picture in some way. I have argued that what is really at issue is the fact that non-attributable agents are alienated from the motivational states that move them into action. And here, it is the coordinated symmetry between who the agent is and what she does and the fact that her action comes about due to a mechanism that plays a part in ensuring that coordinated symmetry that ensures non-alienation. Think about it this way: we're not frightened by the notion of failing to actively pick which motivation to act on and acting anyway—we do so all the time when we let ourselves run on autopilot. What's frightening is the prospect of being autopiloted in directions that have nothing to do with our own perspective of our interests—the motivations that we would give weight to were we to reflect.

¹⁵ Other factors may differentiate the wholly and partially willing in terms of the nature of the response that will be appropriate. For example, if the partially willing addict's taking drugs in this case meets the remaining criteria for blameworthiness, this kind of agent may be less blameworthy than a wholly willing addict, but she is nevertheless an appropriate target of blame.

the right result. For example, imagine a woman whose conflicting first-order motivations are the motivation to take heroin and the motivation to tend to her child, and then imagine her being such that if she were to deliberate about what to do, she would give some weight to each motivation. “On the one hand” she might think to herself, “I want to act on my desire to tend to my child, but on the other hand I do really love heroin a lot and maybe the fun of taking heroin is more important to me than my child’s wellbeing...well, I guess my child’s wellbeing edges out my desire for heroin, so I guess that’s what I’ll do?” In this case, if the woman ends up taking heroin due to her desire to have fun, despite the outcome of her actual or hypothetical deliberation, it seems that she is attributionally-responsible for her action.

However, consider another kind of case that falls somewhere in between the paradigmatic cases of the willing and unwilling addict: the begrudgingly willing addict. Michael Bratman describes the case as involving an addict who, “since [he] is confident that his desire for drugs will soon so overpower him as to prevent him from acting intentionally, and since the struggle to remain drug-free is extremely painful... decides to cease resisting his desire, and to take the steps necessary for satisfying it.” “To be sure,” Bratman writes, “he would rather not perform **an** act of drug-taking. Nonetheless, given his options, he would rather perform *this particular* drug-taking act” (Bratman 1996). Such an addict does minimally endorse acting on his desire to take drugs, and so would count as attributionally-responsible for his action on the view currently under consideration, yet this doesn’t seem like the right result.

To elaborate, the view apparently entails that if the agent were to wholeheartedly resist by not endorsing a desire to take the drug despite inevitably having his will being taken over by the urge, his resultant action would not be attributable. However, if he were to wisely recognize that failure would be the inevitable outcome of his resistance and get it over with more quickly by acting on the desire to take the drug to get rid of the urge faster, his action would be attributable. But it seems that the wrong thing makes the difference in these two cases; the action’s connection to the agent shouldn’t be determined by whether or not the agent decides to give in to an overwhelming impulse, but rather, by whether or not his desire is the kind of thing he’d wanted to be motivated by in the first place. The begrudging addict’s action is still caused purely by neurological noise, no matter how long the agent refrains from acting on it.

People whose psychological make-up has this structure are not limited to hypothetical addicts. Recent research into Tourette syndrome reveals that the initiation of ticcing behavior often comes about in a similar way.¹⁶ Once thought to be akin to involuntary twitches like muscle spasms, Tourettic tics are now acknowledged to be intentional movements consciously undertaken in response to premonitory urges, which are experienced by agents as alien. More than 90% of people with Tourette syndrome report that their tics are “voluntary” in the sense that they believe they take an active (though subservient) role in the action’s coming about (Leckman et. al, 1999). The reported phenomenology echoes the neurophysiological findings. Ordinary action involves signals being sent from an agent’s frontal lobe, the site of considered judgments about what to

¹⁶ Schroeder (2005) argues for the importance of the integration of this research into our theorizing about responsibility.

do, to the motor cortex, the initiator of action. But in the brains of people with Tourette syndrome, disordered neural connections between the basal ganglia and motor cortex cause errant signals to encourage the initiation of simple body movements (including the utterance of words) that build up as a sort of mental pressure over time, making the sufferer more and more uncomfortable until she chooses to act on the urge. People with Tourette syndrome can, with some difficulty, factor in considered judgment to postpone the discharging of these urges, but ultimately usually choose to give in to the urges as a way of alleviating the pressure.

Compare the person who utters slurs for the pure joy of harming minorities to the person with Tourette syndrome who utters slurs due to wanting to relieve the unbearable pressure of a strong premonitory urge to tic. The former, unlike the latter, seems to be agentially involved in the right sort of way with the motivation of her action such that her action represents something about her, while the latter's action does not seem to be representative at all. According to the kinds of endorsement views under consideration so far, however, most agents' tics would be attributable, since it seems that people with Tourette syndrome do give weight to the desire to act on their urges, which comes from their desires to rid themselves of the uncomfortable urges.

Something seems to be going awry in the psychology of the begrudging addict and the person with Tourette syndrome such that the relationships they have to their effective motivating desires do not grant the desires the usual authority to speak for them. Here is what I think goes awry. There are oftentimes two functions of acting on a desire: acting on a desire to φ brings it about so that you are φ -ing, but it also in many cases gets rid of your desire to φ . For example, if you want to go to the library, and then you successfully act on this desire, you will be at the library and thus no longer have an occurrent desire to go to the library. In most cases in which we endorse a desire to φ we do so because we want the satisfaction of the desire to φ , not merely because we want to no longer want to φ . In cases in which we endorse acting on a desire because we approve of doing so, we do so because we have some aim that acting on our desire satisfies other than its elimination through our action. When you 'endorse' giving in to such an urge due solely to its power, the force the urge exerts on you is no less purely mechanistic than when it overpowers your endorsement. To distinguish, I will call the kinds of endorsements that are not merely due to wanting to rid oneself of one's motivation with no further aim "approvals." So, it is approval rather than just any kind of endorsement that is relevant to attributability.¹⁷

¹⁷ Some theorists believe that all addiction and compulsion functions similarly to the way it's described in these cases; that, in a sense, there are no truly unwilling addicts since addiction affects one's judgment about what one should do in this kind of way. See, for example Buss (2012). Most Deep Self theorists, by way of contrast, take it as given that compulsion at least often involves an agent's deep self mental states being overpowered by a rogue urge. But this dispute needn't be settled here. It seems that our view should tell us that irrespective of the prevalence of cases of its type, if there were a case of addiction in which an unapproved desire simply overpowered one's approved desires, the agent's action would count as non-attributable. A related issue is that there may be ego-syntonic cases that we are inclined to call compulsion, in which an agent endorses acting her desire to say, wash her hands 5 times in a row, because she earnestly believes that doing so is an important part of hygiene. These cases seem to me to be cases where an agent's action is self-expressive and therefore she is attributionally-responsible for it, but given her episodic circumstances she is exempt from blame.

Looking at what an agent approves of under more fine-grained descriptions can also shed light on some tricky cases in which an agent is attributionally-responsible for her action under a certain description but not another. Consider a case in which a person appears to act in order to eliminate her desire, but does so to further some aim she has in eliminating it: for example, a person with Tourette syndrome might have an unbearable urge to say something offensive, and be faced with the choice to discharge the urge now, when only one person would be harmed by his doing so, or in 5 minutes, when two people would be harmed. Such a person *does* approve of “saying something offensive now rather than later”, but does not approve of “saying something offensive rather than not saying something offensive.” This tracks the fact that, intuitively, she is praiseworthy for acting now rather than later but not blameworthy for saying something offensive full stop.¹⁸

With all the pieces in place now, we should be in a place to formulate an attractive view of attributional-responsibility. To “minimally approve” of acting in a certain way is to approve of it in just (at least) a hypothetical and partial sense. To be attributionally-responsible for φ -ing, you must be such that you would minimally approve of acting on your effective motivation, and the psychological forces that make it so that you do must play some role in bringing it about that you act in the way that you do. What role, more precisely, should the fact that you would minimally approve play, though? Minimally approving should be non-coincidentally related to why an agent acts. But given the fact that minimal approval is hypothetical, we can’t make this a simple causal connection. What matters is that the chain of mental states that lead an agent to action along with her other psychological features ensure that makes sure that she minimally approves. But in order to rule out fluky cases, such as cases in which they do so by causing her to hit her head in such a way that she becomes such that she minimally approves, we need to make the account dispositional.¹⁹

We can formulate the account now as follows:

THE MINIMAL APPROVAL ACCOUNT OF ATTRIBUTABILITY: An agent is attributionally-responsible for φ ing-at- t ²⁰ iff the actual sequence of mental states involved in the production

¹⁸ It’s worth mentioning the complex relationship between exemption and the fact that many prime examples of agents who are exempt have some sort of disability. Lest someone worry that the very fact that the agents in my examples have named disabling conditions illicitly drives intuitions that they ought to be exempt, it should be noted that an agent need not have a disability to be exempt. The premonitory urge to tic can safely be replaced in the examples with the more familiar urge to scratch an itch, and the view still seems to return the right results. That said, acting in ways that one does not minimally approve of in circumstances when this is rare might go some way towards explaining *why* having the disability might be disabling (and this can be understood on either a medical or a social model of disability). (Thanks to an anonymous Oxford reviewer for pressing me to speak to this important issue.)

¹⁹ We also want to rule in cases in which the only reason she does not minimally approve is due to a similar sort of fluke. To neglect this point would be to run afoul of the Conditional Fallacy. For discussion see Bonevac, Dever, and Sosa (2006).

²⁰ “ φ -ing” should be understood here as standing for an action, and the account should be taken to cover only attributional-responsibility for *actions*. Attributional-responsibility for omissions and for consequences is, on my view, derivative on attributional-responsibility for actions, though I leave the details of the exact bridge principle(s) for another day.

of her action, together with her other mental states, makes it the case that at t , if she were to reflect on her desire to φ at t , she would be sufficiently likely to want to act on her desire to φ at t , with some further aim in doing so other than merely eliminating this desire.²¹

II. ADVANTAGES OF THE VIEW

A. *Capacity Without Process*

We tend to think that human agents' actions can put them on the hook for what they have done in a way that non-agents', like non-human animals' behavior does not. We think the kinds of moral responsibility responses that are appropriate for persons go above and beyond the ways it is reasonable to respond to a dog. Planning, endorsing, and valuing versions of Deep Self views generally aim to be able to make this distinction by positing some special kind of capacity that agents actively exercise when they undertake actions for which they are attributionally-responsible. For example, on the endorsing view, an agent exercises her capacity to choose which action to initiate by picking amongst her first-order desires.

But this feature of these views is also problematic, because the idea that we as agents go through some sort of special mental process each time we act in an attributable way is implausible.²² The problem is that Frankfurtian descriptions of agents selecting which first-order desire to act upon, or Bratmanian descriptions of agents simultaneously acting and making choices about how to settle conflicts amongst desires in the future, just do not seem to be what we ordinarily do as agents in everyday life. Our conduct for which we are often rightly held morally responsible is sometimes spontaneous, initiated by subconscious motivation, out of character, or brought about in a fit of emotion. Furthermore, as the results of numerous social psychology studies appear to show us, we sometimes lack reflective access to some of the motivational influences on our actions, perhaps in ways that would implausibly preclude us from being attributionally-responsible for a large range of actions given the conditions of these more agentially demanding Deep Self views.

On the Minimal Approval view, though, action brought about by subconscious processes can still meet the requirements for attributability. The process that causes action needs to guarantee that the action is in line with what the agent to some degree wants for herself, which requires the agent to be the sort of creature who has the capacity to form higher-order desires. But that capacity need not be exercised in the form of actual reflection, thus avoiding the charge that traditional Deep Self views face that in ruling out animal action, they rule out too much. So a significant advantage of the Minimal Approval view is that it can preserve the distinction between the way agents and non-agents reveal themselves through action due to their having special capacities

²¹ The further aim will often, but need not necessarily be, aimed at good the desire seeks. An agent is also attributionally-responsible, for example, for acting on an urge to hit someone if she does so merely because she wants to know what it would feel like to act on an urge to eliminate it rather than to be aimed at the good the state seeks. (Thanks to Liz Harman for pressing me to revise my view to account for these cases.)

²² See, for example, Arpaly (2002), Smith (2005), Arpaly (2006), and Buss (2012).

while simultaneously allowing that agents may not actually engage in such methodical deliberative processes when they act in ways that they can be held attributionally-responsible for.²³

B. Criterion Operates Independently of the Type of Mental State that Causes Action

A further advantage of the Minimal Approval view is that, unlike some Deep Self accounts, it is compatible with attributable actions being caused by mental states of any type, just so long as they meet the requirement of being appropriately related to hypothetical approval states. This agnosticism about what the actual process of action production looks like is an advantage for at least two reasons.

First, unlike several Deep Self views, accepting the view does not require accepting controversial positions in moral psychology. For example, the valuing view, as advocated by Gary Watson assumes that normative judgments can motivate agents in a way that is distinct from how ordinary desires do, and the caring views of Chandra Sripada and David Shoemaker rely on accepting that there exist complex states or dispositions that we can identify as caring states, which are distinct from mere desires and play a central role in action production. In contrast, the Minimal Approval view is compatible with each of these pictures of action production, but its proponents can remain agnostic about which sorts of states have the ability to motivate. It is even consistent with extremely minimal theories of action production including simple forms of Humean psychology, which might make the view attractive to those who are averse to more traditional Deep Self views due to the more complicated systems of action-production that they posit.

But there is a further advantage to the fact that the Minimal Approval view does not posit that any particular kind of mental state must be involved in the causal chain in order for an action to be attributable: certain mental state types may sometimes produce attributable action and sometimes produce non-attributable action. In order to illustrate this point, I want to focus on a class of actions to which Deep Self theorists have perhaps paid insufficient attention: actions in which agents act directly “out of” emotions. Emotions are often thought to be partly constituted by motivational states or, at the very least, they are generally thought to have some unmediated influence on motivation. This accords with the common-sense view that we can “strike someone out of anger” or “hide out of embarrassment.” Such actions often characteristically do not align with our plans, values, cares, or second-order volitions concerning what we think the most preferable thing to do is in a given situation. Imagine, for example, an anti-retributivist who nevertheless is swept up in a wave of vengeful anger, or an ethically non-monogamous person who has disavowed the appropriateness of jealousy being nevertheless moved to action by it. While these actions fail to align with Deep Self states, they nevertheless seem to be the sorts of things for which one can be attributionally-responsible.

²³ While the Minimal Approval view has this advantage, the minimalist nature of the view might seem to run a different risk, namely, being unable to plausibly explain why agents can be attributionally-responsible in a way that most non-human animals cannot. However, the idea that humans are unique among animals in being able to have higher-order thoughts has a rich history of support in the literature on consciousness. See Carruthers (2016) for discussion.

On the other hand, if a theory were to hold that a person is attributionally-responsible for *any* action done out of emotion, it would not be viable. Actions caused by psychological and neurological disorders that we intuitively tend to think exculpate an agent from attributional-responsibility can, it seems, cause action by impacting the agent's emotional state such that she "acts out of" a given emotion. If this is right, a theory of attributional-responsibility should have a way of distinguishing between cases of acting out of an emotion that involve one's agency in the right sort of way and ones that circumvent agency.

One lesser-known disorder illustrates the importance of drawing such a distinction. Misophonia is a neurobehavioral syndrome in which certain ordinary human-produced repetitive sounds, (such as the sounds of others chewing, sniffing, or clearing their throats), trigger reactions of anger, disgust, and fear in otherwise psychologically healthy individuals.²⁴ While research on misophonia is in its infancy, it is hypothesized that the cause of such reactions is extra-connectivity between a set of emotional processing centers of the brain and the anterior insular cortex, the site of interoception (the ability to sense what is happening to one's own body) in the brain.²⁵ Due to this over-connectivity, ordinary sounds cause these sufferers to react as if these innocuous sounds are threats, setting off fight-or-flight reactions. When misophonia sufferers are in "fight" mode, their anger is not just an expression of being overwhelmed, but rather, tends to take the form of a directed expression of anger and disgust towards the source of the offending sound. To be clear, this is not just anger towards the person for making a sound that they know bothers the sufferer, as anger can be just as strong towards those making sounds who do not realize their sounds are upsetting to the sufferer. Crucially, at the very same moment in time that she acts out of anger, a person with misophonia is able to acknowledge that it makes no sense to be angry and that, for example, making sounds while chewing is entirely innocuous. Due to these irrational, embarrassing, and inescapable responses, people with misophonia often live increasingly reclusive lives as the disorder progresses in order to try to avoid both sounds and accidentally lashing out at those who they know have done nothing wrong.

Consider the following pair of cases:

MANNERS MARY: Manners Mary was taught as a child to always chew with her mouth closed and greatly appreciated the value of the lesson. Following her parents, she grew up believing that a decline in manners in society was the root of much evil and that it is deplorable that some people chew with their mouths open. In her adulthood, she has come to see this as a bit overblown, but she has retained the sense that it's bad form to chew with an open mouth as well as an accompanying sense of disgust when she sees others behaving with such poor manners. At an important dinner party she notices her fellow guests chewing with open mouths, and thinks to herself that someone ought to tell them to stop, and perhaps if no one else does, she should be the one. However, she knows that these guests would only be offended and would not change their ways if she were to men-

²⁴ See Braut et. al (2018) for a cross-disciplinary review of the research on misophonia.

²⁵ See Kumar et. al (2017) , Edelstein et. al. (2017)

tion their behavior, and so decides that this would probably be a bad time to say something. However, she fails to hold her tongue, gets increasingly angry, and yells out “chew with your mouth closed!” despite knowing that everyone will only be offended and not change their ways.

MISOPHONIA MARY: Misophonia Mary does not care about manners in the slightest and makes no effort to chew with her mouth closed. However, she suffers from misophonia, which makes her inexplicably angry when she hears people making chewing sounds. Even when in the throes of an episode of misophonia she recognizes that there is nothing wrong or bad in any way about eating with one’s mouth open, yet due to errant signals in her brain that trigger a fight-or-flight reaction, Mary feels compelled to flee or else lash out at those making the sounds. With nowhere to flee to at a dinner party, out of anger Mary yells out “chew with your mouth closed!” despite knowing that everyone will only be offended and not change their ways.

Intuitively, it seems we should hold Manners Mary attributionally-responsible but not Misophonia Mary, though they are both most directly motivated by their anger. This is some indication that our view should be consistent with the fact that acting out of anger is neither sufficient for attributional-responsibility nor disqualifying for it.

But because both agents’ actions are motivated by anger and not suitably related to their plans, endorsements, judgments, or possibly even cares, traditional Deep Self views will have a difficult time explaining why Manners Mary’s action is attributable and thus licenses a different response than Misophonia Mary’s.

The Minimal Approval, by contrast, is well-suited to explain the contrast. If we were, at the time of action, to ask Manners Mary to consider her motivation to yell out at the guests chewing with their mouths open she would give *some* weight to that option. After all, she thinks it is somewhat important that such ill-mannered behavior not go wholly ignored.²⁶ But it does not seem appropriate to hold Misophonia Mary attributionally-responsible for her yelling, and the Minimal Approval view shows how her motivation stands outside of her agency. If we were to ask her whether or not she would like to be motivated to some degree by the desire to lash out at people chewing she would say, even in a moment of her anger, that she has no desire to be so moved. The only reason she might give any weight to the desire to lash out would be to relieve

²⁶ Importantly, even if Mary were merely taken over by a fleeting wave of disgust or in a petty mood, the relevant counterfactual worlds by which we should evaluate whether or not she would minimally approve of yelling out are the ones in which her fleeting wave of disgust or petty mood are held fixed. It must be the case that she reflects in those worlds, but that she does not do so coolly if her psychology at the time of action is not similarly cool. (Relatedly, the worlds of interest are ones in which Mary considers her first-order desires in the actual world to act that may pull in different directions, but not ones in which Mary is unusually well-informed about outcomes or alternatives.) This differentiates the view from the caring-based deep self views, since, for example Wave-of-disgust Mary who does not really care about impolite chewing seems as though she would minimally approve of acting as she does while her action while her motivation would fail to be sufficiently related to what she cares about to speak for her on a caring-based deep self view.

the psychological pressure of not saying anything, caused by her involuntary fight-or-flight reaction, and thus she would fail to meet the conditions of the Minimal Approval view.

This pair of cases helps illustrate the fact that the Minimal Approval view can hold that agents who act out of emotions are often attributionally-responsible for their actions while leaving room for the possibility that emotional motivation may factor prominently in action caused by neurological noise for which we should not hold agents attributionally-responsible. The set of cases in which agents act out of emotions helps illustrate the broader point that in having a set of criteria for attributability that does not require any particular mental state *type* to feature in the action-causing sequence it has better flexibility for handling some of the nuances of attributability and neurological dysfunction.²⁷

IV. CONCLUSION

I have advanced a new account of the necessary and sufficient conditions for attributability, understood as that which gives actions the ability to speak for agents or connects acts to persons' agency. I have proposed that the relevant condition is that the agent endorse her action in a very minimal sense: the endorsement may be merely hypothetical, and it may be merely partial. The endorsement, I argued, must be a desire to act on a certain desire, which must serve some further aim than merely getting rid of one's desire. I highlighted several advantages of this view over traditional Deep Self accounts. Along the way, I showed that this account appropriately handles cases of compulsion, addiction, weakness of will, Tourette's syndrome, and misophonia.

Despite these considerable advantages, additional work for the Minimal Approval theorist, of course, remains. She will have to give an account of desire on which partial desires to satisfy first-order desires to ϕ are clearly demarcated from first-order desires to ϕ . She will also bear the burden of proof to show that there are facts of the matter about what agents would counterfactually endorse. And furthermore, she must show that, despite the fact that the states are counterfactual, they are grounded in the kinds of psychological dispositions that do tell us something valuable about what agents are *actually* like.

Though I will not be able to provide a full defense of this claim here, I believe that, with some exceptions, agents who meet the conditions on attributability according to caring, valuing, planning, and endorsing Deep Self views will largely also meet the conditions for minimal approval. I take the action-generating processes that traditional Deep Self views target to often be sufficient for attributability, but believe none are necessary, as many other action-generating processes may also meet the conditions of minimal approval. So in one sense, the Minimal Approval

²⁷ One further advantage of the view is worth noting. Moving the criterion to partial endorsement perhaps helps avoid one version of a popular criticism to Frankfurt's hierarchical view. Frankfurt speaks of endorsing one's desire as an authorizing process, which raises the question: shouldn't an agent also have to endorse her *endorsement* to grant *it* the requisite authority to speak for her? On the view I am proposing, in contrast, the fact that the agential process secures the bare presence of some reflective identification suffices to make the action generated by the process able to speak for the agent. Minimal Approval theorists may also take refuge from another worry, that even partial second-order endorsements may be compulsive, by showing how these endorsements will fall short of being genuine states of *approval* in the sense I've developed.

view can be seen as more austere, as it is robust across a wider variety of different causal stories.²⁸

On the other hand, by design, meeting the requirements of the Minimal Approval view is not particularly agentially demanding, and the resulting picture of the self is rather fragmentary, so nothing as exceptionally deep about an agent's character or practical identity can properly be said to be revealed through her attributable actions. Because of this, the Minimal Approval theorist is tasked with ensuring that the sense of attributability she articulates really can ground an account of attributional-responsibility that is central and important to our moral responsibility practices.²⁹

Furthermore, for all that the view differs from more traditional Deep Self theories, it still shares some of their burdens. Most pressingly, the Minimal Approval theorist will have to have some response to familiar counterexamples about manipulated agents whose attributability-granting mental states are implanted by neuroscientists rather than acquired in the normal fashion.^{30,31}

While these obstacles remain, they are not insurmountable. I hope to have shown these obstacles to be worth facing, as I believe that the Minimal Approval view has several features that make it a viable contending account of the conditions for attributability more than worthy of serious consideration.

²⁸ In this way, the view in some ways resembles the class of views Matt Talbert calls New Attributionist, which he takes to include his own views as well as the views of T.M. Scanlon, Angela Smith, and George Sher (Talbert 2016). While these views differ in many ways, they are each relatively austere in that they posit both that agents are attributionally-responsible for a much broader range of behavior than traditional Deep Self theorists do and that attributability is sufficient for holding an agent fully accountable.

²⁹ But note that while certain paths from attributability to blame may be ruled out by ceding these more complex Deep Self agent architectures, there is independent reason to believe that the agential traits to which *aretaic* responses are warranted are finely individuated. For example, evidence from Situationism in psychology that casts doubt on certain grander notions of the Self may be compatible with more fragmentary understandings of character.

³⁰ This problem may even be especially pressing for the Minimal Approval view (although it shares this feature with Frankfurt's second-order volition view), since it does not require that any states that guarantee cross-temporal agential connections need take part in the production of attributable action.

³¹ See Matheson [forthcoming] for a recent overview of the state of this debate.

References

- Arpaly, Nomy. (2002). *Unprincipled Virtue: An Inquiry Into Moral Agency*. (Oxford: Oxford University Press).
- Bonevac, Daniel, Dever, Josh and David Sosa. (2006). "The Conditional Fallacy." *Philosophical Review* 115 (3):273-316.
- 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.
- Braut, Jennifer J. (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.
- Carruthers, Peter. (2016). "Higher-Order Theories of Consciousness", *The Stanford Encyclopedia of Philosophy*, Edward N. Zalta (ed.)
- Doris, John. (2002). *Lack of Character* (Cambridge: Cambridge University Press).
- Edelstein, Miren et al. (2013). "Misophonia: Physiological Investigations and Case Descriptions." *Frontiers in Human Neuroscience* 7, 296.
- Fischer, John Martin. (2012a). "Responsibility and Autonomy: The Problem of Mission Creep." *Philosophical Issues* 22(1), 165-184.
- Fischer, John Martin. (2012b). "Semicompatibilism and Its Rivals." *Journal of Ethics* 16 (2), 117-143.
- Frankfurt, Harry. (1969). Alternate Possibilities and Moral Responsibility. *Journal of Philosophy* 66 (23):829.
- Frankfurt, Harry. (1982). The Importance of What We Care About. *Synthese* 53(2), 257- 272.
- Frankfurt, Harry G. (1988). *The Importance of What We Care About: Philosophical Essays*. Cambridge University Press.
- Frankfurt, Harry. (1971). Freedom of the Will and the Concept of a Person. *Journal of Philosophy* 68(1), 5-20.
- Gorman, August. Unpublished Manuscript. "How Should a Deep Self Theorist Handle Weakness of Will?" Georgetown University, Department of Philosophy.
- Haji, Ishtiyaque. (1998). *Moral Appraisability: Puzzles, Proposals, and Perplexities*. (Oxford: Oxford University Press.)
- Haji, Ishtiyaque. (2002). "Compatibilist Views of Freedom and Responsibility." In Robert H. Kane (ed.), *The Oxford Handbook of Free Will*. (Oxford: Oxford University Press.)
- Jaworska, Agnieszka. (2015). "Identificationist Views," In Meghan Griffith, Neil Levy, and Kevin Timpe (eds.), *Routledge Companion to Free Will*, (Abingdon: Routledge).
- Kumar, Sukhbinder et al. "The Brain Basis for Misophonia." *Current Biology* 27.4 (2017): 527–533. *PMC*. Web. 14 Sept. 2017.

- Leckman, James F. and Donald J. Cohen. (1999). *Tourette's Syndrome—Tics, Obsessions, Compulsions: Developmental Psychopathology and Clinical Care* (Hoboken: Wiley-Blackwell), 27.
- Matheson, Benjamin. (Forthcoming). "The Threat from Manipulation Arguments." *American Philosophical Quarterly*.
- McKenna, Michael (2008). Frankfurt's Argument Against Alternative Possibilities: Looking beyond the examples. *Noûs* 42 (4): 770-793.
- McKenna, Michael. (2011). "Contemporary Compatibilism: Mesh Theories and Reasons-Responsive Theories." In R. Kane, ed., 2011, *Oxford Handbook of Free Will*, 2nd ed. (New York: Oxford University Press), 175-98.
- McKenna, Michael and Chad Van Schoelandt. (2015). "Crossing a Mesh Theory with a Reasons-Responsive Theory." In A. Buckareff, C. Moya, and S. Rosell, *Agency and Responsibility* (Basingstoke: Palgrave Macmillan).
- Mele, Alfred R. (2006). *Free Will and Luck*. (Oxford: Oxford University Press.)
- Sartorio, Carolina (2016). A Partial Defense of the Actual-Sequence Model of Freedom. *The Journal of Ethics* 20 (1-3):107-120.
- Schroeder, Timothy. (2005). Moral Responsibility and Tourette Syndrome. *Philosophy and Phenomenological Research* 71(1), 106–123.
- Shoemaker, David. (2011). "Attributability, Answerability, and Accountability: Toward a Wider Theory of Moral Responsibility." *Ethics* 121(3), 602-632.
- Shoemaker, David. (2003). "Caring, Identification, and Agency." *Ethics* 114 (1):88-118.
- Shoemaker, David. (2015). "Ecumenical Attributability" in *The Nature of Moral Responsibility* ed. Randolph Clarke, Michael McKenna, and Angela Smith. (Oxford: Oxford University Press.)
- Silverstein, Matthew (2017). Ethics and Practical Reasoning. *Ethics* 127 (2):353 - 382.
- Smith, Angela M. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life." *Ethics* 115 (2):236-271.
- Sripada, Chandra. (2015). "Self-expression: A Deep Self Theory of Moral Responsibility." *Philosophical Studies* 173(5), 1203-1232.
- Strabbing, Jada Twedt (2016). "Attributability, Weakness of Will, and the Importance of Just Having the Capacity." *Philosophical Studies* 173(2), 289-307.
- Talbert, Matthew (2016). *Moral Responsibility: An Introduction*. (Cambridge: Polity).
- Vihvelin, Kadri. (1994). "Are Drug Addicts Unfree?" In S. Luper-Foy and C. Brown (eds.) *Drugs, Morality and the Law* (New York: Garland), 51-78.
- Velleman, David J. (1992). "What Happens When Someone Acts?" *Mind* 101(403), 461- 481.
- Watson, Gary (1975). "Free Agency." *Journal of Philosophy* 72 (April), 205-20.
- Watson, Gary (1996). "Two Faces of Responsibility." *Philosophical Topics* 24(2), 227-248.