

Blameworthiness and Constitutive Control

ABSTRACT: According to “voluntarists,” voluntary control is a necessary precondition on being blameworthy. According to “non-voluntarists,” it isn’t. I argue here that we ought to take seriously a type of voluntary control that both camps have tended to overlook. In addition to “direct” control over our behavior, and “indirect” control over some of the consequences of our behavior, we also possess “constitutive” control: the capacity to govern some of our attitudes and character traits by making choices about what to do that constitute those attitudes and traits. Taking this sort of control seriously, I argue, ultimately tips the scale towards voluntarism. First, I address a non-voluntarist case in which an agent is putatively made blameworthy by the reasons for which she acts, even though the particular reasons for which she acts aren’t up to her. I argue that this case looks compelling only if we overlook constitutive control, and thereby miss how the agent’s motivating reasons are under her voluntary control even though non-voluntarists think they are not. I then use the notion of constitutive control to diffuse some of the best putative counterexamples to voluntarism: cases in which subjects are blameworthy either for caring inadequately about others or for wishing them ill.

I. Voluntarism and Non-voluntarism

Philosophers who write about blameworthiness, or, more generally, about the type of moral responsibility that renders one a legitimate target of blame, are divided over voluntary control. Specifically, they are divided over the commonsense type of control that one might invoke in saying, “I can’t control whether it’s going to rain today, but I can control whether I bring an umbrella”: our capacity to make a difference to what happens via our choices about what to do.¹ According to “voluntarists,” people can only be blameworthy for things under their voluntary control. According to “non-voluntarists,” people can be blameworthy for things that are not under their voluntary control, such as the non-voluntary possession of objectionable other-regarding attitudes.²

¹ For an elaboration on this commonsense notion of “control,” see Adams (1985, 8-10). Whether voluntary control is compatible with determinism is a further matter that we’ll leave aside.

² Some voluntarists include Audi (1991), Carlsson (2017), Fischer and Ravizza (1998), Fischer and Tognazzini (2012), Levy (2005), Moody-Adams (1990), Rosen (2015), and Wallace (1994). Some non-voluntarists include Adams (1985), Graham (2014), Hume (1740/1978), Hieronymi (2008, 2014), Scanlon (2008, 2013), Sher (2006), Shoemaker (2003), Smith (2005, 2008), and Wolf (2011). Philosophers such as Hieronymi (2008, 2014) and

This dispute is sometimes framed as about whether blameworthiness is ultimately grounded in one's *character* or in one's *behavior* (where "behavior" is broadly construed to include any events voluntarily controllable via choice – including whether we make mere attempts and preventable omissions).³ When the debate is thus cast, voluntarists claim that people are only ever blameworthy because of what they *do* (broadly construed), and are only ever blameworthy *for* something because it is either what they have done or a (reasonably foreseeable) consequence thereof. Non-voluntarists, on the other hand, contend that people are sometimes blameworthy simply in virtue of, and for, who they *are* – even if who they are isn't up to them to choose. In fact, non-voluntarists believe, people are only ever blameworthy for what they *do* derivatively, when and because their choice-making reflects a deeper ill constitution.

A typical exchange between a voluntarist and non-voluntarist may proceed thus: The former will claim that she holds the initial advantage on the grounds that it is a pre-theoretical platitude that agents are only blameworthy for things under their voluntary control. There is an "intuitively powerful distinction between *bad* agents and *blameworthy* agents," (Levy 2005, 2) between "a *faulty* attitude (character, act or omission) and one for which an agent is *at fault*" (Levy 2005, 5). The non-voluntarist will then attempt to chisel away at these convictions by pointing out that there are many cases in which subjects are blameworthy for things that they have not chosen to bring about: What about the man who believes that women naturally have less aptitude for math? A person doesn't simply choose whether or not to form such a belief, yet he is plausibly blameworthy for holding it. The voluntarist will reply (predictably) that an

McHugh (2017) who think that reasons-responsiveness is a non-voluntary type of "control" and that this is the only type of "control" required for blameworthiness count as *non*-voluntarists.

³ See, for example, Graham (2014), Hieronymi (2014), and Hume (1740/1978).

agent is only ever blameworthy for something that he did not *directly* bring about, because it was under his *indirect* control: it was the (reasonably foreseeable) consequence of one of his choices or omissions.⁴ Incorrect and harmful aptitude assessments are often the predictable results of culpable failures to attend to easily accessible evidence. To which the non-voluntarist will in turn retort that the method of “tracing” what an agent is blameworthy for back to some voluntary choice or omission doesn’t seem to work in all cases. Sometimes a person’s development of bad character traits may have been unforeseeable and unpreventable, and yet she can be blameworthy for them nonetheless.⁵

At this point the voluntarist might suggest that she and the non-voluntarist just have in mind different forms of negative assessment. Of course a person’s non-voluntary flaws can render her worthy of adverse evaluation, even adverse “aretaic evaluation” that redounds deeply to her identity as an agent.⁶ But what the voluntarist is talking about are the conditions under which people are worthy not just of negative character assessment, but rather of being *held* responsible. More specifically, she might add, she is talking about notion of blame that is most central to our practices of holding one another to account: about the conditions under which agents are worthy of responses that include the reactive attitudes of resentment, indignation, and (in the self-directed case) guilt. But the non-voluntarist will reply that she too is talking about the conditions under which agents are worthy of being *held* responsible. She is talking about “full blooded notions of moral appraisal, blame, and criticism” and her claim is that voluntary control is not a necessary prerequisite for being worthy of this sort of response

⁴ See, for example, Fischer and Tognazinni (2012).

⁵ See, for example, Adams (1985, 12-13) and Sher (2006, 52-55).

⁶ Watson (1996) distinguishes this sort of evaluation from what he calls “accountability blame,” and suggests that voluntary control is a necessary precondition for being worthy of the latter but not the former. For another attempt to distinguish accountability blame from other types of blame, see Shoemaker (2011).

(Smith 2008, 380).⁷ After all, she'll maintain, characteristics such as failing to value others sufficiently, wishing them harm, or delighting in their pain *do* intuitively make subjects worthy of resentment, indignation, and guilt, and do so even when there is no instance of faulty choice or omission to which such vices can be sourced.⁸

Voluntarists and non-voluntarists alike accept that we have direct control over our behavior via choice. And both camps recognize that we have indirect control over the foreseeable results of our behavior, which can include “managerial” control over our own attitudes and character traits. My aim in this article is to insist that we take seriously a third type of voluntary control. In addition to direct and indirect control, we also have what I will call “constitutive” control: the ability to control some aspects of our character by making choices about what to do that constitute them. That we have this sort of voluntary control should be unsurprising. The thought that *who we are* can be constituted by *what we do* is new neither to philosophy nor to “folk” wisdom. (Lest you’ve forgotten: you are what you eat.) Yet it has not been afforded much deference in discussions of blameworthiness. I aim to alter this state of affairs for two

⁷ Indeed, *both* voluntarists and non-voluntarists have claimed to be talking about the same thing, and both have claimed to be talking about the only sort of blame (and blameworthiness) that there is (Levy 2005, 4, and Smith 2012, 576).

⁸ A complication in this debate is that, although they both claim to be theorizing about the notion of blame central to practices of holding responsible, I think the voluntarist and non-voluntarist probably *do* have somewhat differing conceptions of what this amounts to. For instance, Smith (2012) *concedes* that she thinks of blame in terms of Shoemaker’s (2011) notion of “answerability” as opposed to “accountability.” That is, Smith (2008, 381) thinks that what is distinctive about blame is that it “calls upon the agent to explain or justify her rational activity in some area.” She just also believes that to “address a demand” to an agent in this way *is* to robustly hold her responsible, and, indeed, that this is *the only sort of blame that there is* (Smith 2008, 381 and 2012). But I am inclined to rejoin that “addressing a demand” in Smith’s sense is *not* robust enough to constitute blame as I think of it, and that to appeal merely to this sort of “address” is *not* sufficient to capture what’s distinctive about resentment, indignation, and guilt. As I’ll soon elaborate in the main text, I think that blame is *retributive* in the sense that it involves a readiness to make some target *feel bad*. Yet non-voluntarists and I do appear to be discussing *overlapping* phenomena, and so the non-voluntarists’ arguments still require a response. Indeed, what is particularly worrisome is that non-voluntarists have produced cases in which agents appear to be worthy of blame *as I understand it* even for things that aren’t under their voluntary control (such as when agents don’t care adequately about others or wish others ill). Thus, I’ll here be responding to non-voluntarist arguments interpreted as challenges to the view that voluntary control is a precondition for blameworthiness *in the voluntarists’ “accountability” sense*.

reasons. First, I am concerned that the terms in which these discussions have tended to be cast obscures some of the power we have over who we are. This strikes me as generally problematic, because I think such power merits owning up to. Second, taking constitutive control seriously diffuses some of the best putative counterexamples to voluntarism (or so I will argue) and ultimately tips the scale back towards voluntarism.

I will begin my argument, in Section II, by considering a case that T.M. Scanlon (2008) raises against voluntarism in which an agent is made blameworthy by the *reasons for which* she acts. Scanlon's case presents a challenge to voluntarism, because it looks doubtful whether the agent in it has any control over the particular reasons that motivate her. In Section III, I will take up the challenge, arguing that skepticism about this agent's control over her motivating reasons only seems compelling when we overlook the type of voluntarily control I call "constitutive." Constitutive control explains how her motivating reasons *are* under her voluntary control even though Scanlon thinks they are not. With the notion of constitutive control under our belts, in Section IV, I will use it to disarm some of the strongest putative counterexamples to voluntarism: cases in which agents don't care about others or desire their suffering. In Section V, I will conclude.

One final clarification about what is at issue: I take the question of whether a person is blameworthy to be the question of whether it is fitting to blame her. "Fittingness" is a primitive and *sui generis* normative relation, on which I believe we all have an intuitive grip. It is the normative relation that obtains between emotions and their formal objects, as well as between some actions and certain types of (abstract) objects: between fear and the fearsome, admiration

and the admirable, loathing and the loathsome, and (purportedly) between printing in *The New York Times* and the newsworthy.⁹

Blame, as I'll conceive of it here, does consist in the Strawsonian "reactive attitudes" of resentment, indignation, and (in the self-directed case) guilt – including the cognitive, physical, attentional, and behavioral dispositions and displays that are partially constitutive of such emotions.¹⁰ I believe that an important feature of the blaming emotions is that they have a *retributive* character, and are thus distinguishably more *severe* than other forms of negative assessment. This does not mean that those who experience blaming emotions will always behave with the aim of hurting the target of their emotions, or even express their blame at all, but I do think resentment, indignation, and guilt all involve at least an urge to make their targets feel bad for what they have done or are like, and that they all involve representing their targets as fitting objects of such treatment.¹¹ Moreover, I think this sort of readiness to

⁹For discussions of fittingness and its distinction from other normative notions, see D'Arms and Jacobson (2000a, 2000b) and Howard (2018). Although many philosophers concur that to be blameworthy is to be a fitting target of blame, there are others who prefer to analyze blameworthiness in terms of whether blame is morally fair, or deserved, or felicitously addressed. See Wallace (1994), Pereboom (2014), and MacNamara (2015), respectively. I confess that not all of these alternatives appear to me to be as distinct from one another as some philosophers take them to be. Desert, for instance, just seems to me to be a *species* of fittingness. But there are reasons to resist analyses of blameworthiness in terms of alternative normative notions in any case. For one thing, considerations of theoretical unity speak for analyzing blameworthiness in terms of fitting blame. Understanding blameworthiness in terms of fittingness renders it one amongst *many* evaluative properties analyzable in terms of the response it is fit for. For another, considerations adduced by other philosophers have convinced me that the prospects for extensional adequacy of alternative analyses are not great. For instance Graham (2012, 391-392) argues that it *could* be fair in a *moral* sense of "fairness" to blame individuals who aren't blameworthy, insofar as those people *want* to be blamed. And Rosen (2015, 82) rightly wonders why, if being worthy of blame is a matter of being an apt communicative target, then what matters in determining whether someone is blameworthy for some behavior are that person's capacities *at the time of action* rather than *at the time of blame*.

¹⁰ Sometimes philosophers distinguish between blame and expressions of blame. I think this can be misleading. Blame (and resentment, etc.) can be *unexpressed*, but we should not take this to imply that the behaviors displayed in the throes of emotional episodes are not themselves constituent parts of the emotions when those behaviors do occur. On the contrary, I think yelling (in conjunction with other psychological and physiological effects) is a *way* of blaming (and of resenting). For a similar point, see McKenna (2013, 126).

¹¹ Rosen (2015, 82-84) and Carlsson (2016) both make similar claims. Thinking of blame as involving this sort of *readiness* to sanction is a way of specifying the sense in which both Watson (1996) and Shoemaker (2011, 631) take accountability blame to involve "sanctioning emotions."

sanction cannot be fitting unless at least a very minimal form of sanctioning is itself fitting. Thus, I think we can use our intuitions about whether subjects are fitting targets of emotionally painful sanction (including self-sanction) as a heuristic for determining whether they are blameworthy.¹²

II. Motivating Reasons

An observation often marshaled in favor of non-voluntarism is that subjects can be blameworthy in virtue of the *reasons for which* they do something, even while *what* they do is perfectly permissible. A person can be blameworthy “in the motivation and attitude with which he usually does what he ought to do” (Adams 1985, 5), or for doing “the right thing but for the wrong reason” (Scanlon 2008, 57).¹³ But, non-voluntarists claim, when a person is blameworthy in virtue of her motivating reasons, it is often not the case that whether to be motivated by such reasons was under her voluntary control. Consider:

Rescue to Riches

A person that I hate, and who I would be happy to see die, is drowning. I don't want to save him, but I realize that if he dies right now, then his heir, with whom I am currently locked in a bitter electoral contest, will inherit a large sum of money to spend on her campaign. Thus, in spite of my hatred, I save him – but only because of the political situation (Scanlon 2008, 57).

Scanlon thinks that we should have the intuition that I am blameworthy in this case. But if I am, I can't be blameworthy because of what I *do*, since saving him is surely what I *should* do. (Indeed I would be blameworthy if I *didn't* try to save him.) Thus I must be blameworthy not because of what I do, but rather because I do it for objectionable reasons. However, Scanlon

¹² To say that sanction is fitting is, of course, not to say that it is what is morally or rationally *to be done*. Rather the test for whether an agent is blameworthy will be to check whether sanctioning and readiness to sanction seem intuitively *called for*, in a manner analogous to how a funny joke seems to call for laughter even when laughing is neither morally nor rationally the best thing to do.

¹³ Graham (2015, 395-396) also raises this sort of case.

thinks, I have no control over which reasons motivate my action because I cannot *choose* the specific reasons for which I save him. If this is right, then *Rescue to Riches* shows that agents can be blameworthy for things that are not under their control, namely the reasons for which they perform particular actions.

One may doubt whether I am really blameworthy in this case, but I think it is plausible enough to hold that I am, so let us concede the point. (If you lack the intuition, it helps to imagine that I can barely stomach each stroke I take, and resort to constantly reminding myself of the potential political catastrophe in order to encourage myself to swim forwards.) But is it true that I have no control over my motivating reasons: that I cannot simply choose to act for *other* reasons? It can seem doubtful that I have such a capacity. Insofar as I can choose which *actions* I perform, I can bring it about that I am motivated by one reason, and not another, by choosing to perform an action that I take the first reason, but not the second, to support. But the question presently at issue is not whether I can choose *not* to save the man, but rather whether I can choose *to* save him for reasons *other* than the political situation. And it seems implausible that *this* is something I can choose to do. Consider the following thought-experiment that Scanlon draws from a discussion of Thomson's (1991) on the doctrine of double effect:

The Commander

Suppose you are prime minister, and the commander of the air force describes to you a planned raid, that would be expected to destroy a munitions plant and also kill a certain number of civilians, thereby probably undermining public support for the war. The commander asks you whether you think the raid is morally permissible and you – being a proponent of the doctrine of double effect – reply, “Well, that depends on what your intentions would be in carrying it out. If you intend to kill the civilians then the raid is impermissible, but if their deaths would be merely an unintended foreseeable side effect, it is permissible.”

The case illustrates that there is some oddity in supposing (as proponents of the doctrine of double effect seem to do) that the reasons for which we perform a particular action are up to

us. As Thomson (1991, 293) says in her original discussion, “Can anyone really think that the pilot should decide whether he may drop the bombs by looking inward for the intention with which he would be dropping them if he dropped them?”

Now, perhaps considering one’s intentions during the course of deliberation is not really all that strange.¹⁴ But pressing this reply just misses a deeper concern about choosing one’s reasons that the case of *The Commander* raises. What should really worry me, one might think, is not the oddity of how the pilot may proceed to deliberate, but rather how he is supposed to successfully perform the course of action he alights on. For, suppose that the scene carries on:

The Commander (continued)

The commander asks you whether you think the raid is morally permissible and you reply, “Well, that depends on what your intentions would be in carrying it out. If you intend to kill the civilians then the raid is impermissible, but if their deaths would be merely an unintended foreseeable side effect, it is permissible.” So the commander replies, “Alright, in that case I’ll conduct the raid just in order to destroy the munitions plant,” and goes ahead and bombs the plant.

If this is what happens, *has* the commander bombed the plant in order to destroy enemy weapons? It seems that, in attempting to select destroying enemy weapons as his reason for bombing the plant, the general has unwittingly not acted for that reason at all. Rather he has acted for a different reason altogether: he bombs the plant because he thinks doing so with certain intentions is permissible.

Thus the case for skepticism about choosing our motivating reasons seems quite strong. The problem with the idea that agents can choose the reasons for which they act, or so the reasons-choosing skeptic might suggest, is that the choice to perform some action (A) for some reason

¹⁴ Hanser (2005, 460) offers this response to Thomson (1991).

(R) *itself* needs to be made for a reason. But the reasons to A-for-the-reason-R are generally not the *same* as R, and so being motivated by them is *not* necessarily to be motivated by R. Rather, it is to be motivated by considerations that support *A-ing-for-the-reason-R*. Moreover, so long as being motivated by considerations that support A-ing-for-the-reason-R *doesn't* entail being motivated by R, an agent's choice to A-for-the-reason-R will be ineffective. In trying to make it, he may very well A. But he won't be A-ing for the reason for which is *trying* to choose to A (R). Rather, he will be A-ing for some *other* reason: namely, whatever consideration supports A-ing-for-the-reason-R. Thus, the skeptic will put forward, choosing to act-for-particular-reasons generally results in being motivated by *different* reasons than those that the agent is choosing to act for, and this suggests that, in general, agents cannot choose to act-for-particular-reasons.¹⁵

But if agents can be blameworthy for the reasons for which they perform an action, and yet cannot *choose* the reasons for which they perform that action, then how *can* voluntary control be a necessary precondition on blameworthiness? Perhaps the non-voluntarists are right that it is not.

III. Constitutive Control

¹⁵ I think it is because of Thomson's (1991) thought experiment that Scanlon (2008) briefly considers the possibility that I might control the particular reasons for which I save the man, not by choosing the reasons for which I save him, but rather by bringing about a change in which considerations I see as counting sufficiently in favor of doing so. Perhaps, the thought is, I could make it the case that I save the man for some reason other than the political situation by making it the case that I see that other reason, and not the political situation, as counting sufficiently in favor of saving him. (In fact, in light of Thomson's thought experiment, Scanlon seems to think that this is the *only* way I could bring it about that I save him for one reason and not another (56, 60.)) But Scanlon quickly determines that this strategy won't work either. For, he thinks, I cannot voluntarily decide to see certain considerations as counting (or counting sufficiently) in favor of a particular action. "Deciding in such a case is not choosing," Scanlon says, "because it lacks the relevant element of free play" (60). Thus I cannot choose the particular reasons that motivate me – *either* by choosing what to be motivated by *or* by choosing what to see as sufficient reason.

Scanlon (2008) himself doesn't actually think that an agent's motivating reasons are *never* under his control. In particular, he appears to allow that agents have control over their motivating reasons when the actions they undertake are in pursuit of larger plans. To see why *Rescue to Riches* is *not*, in fact, a counterexample to voluntarism, it is instructive to examine this sort of case:

Rat Man

A man goes to the store to purchase some rat poison because he plans to put the poison in his wife's food and, thereby, kill her.

Are Rat Man's motivating reasons up to him? Scanlon writes:

One thing we can say about a person who is buying poison with the intention of using it to kill his wife is that what he intends to do is impermissible, and that he should abandon that intention (41).

Here, Scanlon suggests that Rat Man *can* control whether he acts for certain reasons, by abandoning his intention to kill his wife. "The larger intention..." Scanlon writes, "suppl[ies] the reason for him to purchase the poison" (41). Thus, presumably Scanlon thinks that if Rat Man abandons his intention, then his plan to kill his wife will no longer be a reason *for which* he purchases the poison. Of course, if Rat Man abandons his plan to kill his wife, then ostensibly he won't purchase the poison at all, since that plan was all that appeared to him to give him reason to, to begin with. So this is a case in which an agent can control what motivates him simply by electing not to act at all.

But Scanlon also appears to allow that agents can control their motivating reasons in a similar manner when *various* reasons for which they might perform an action correspond to different

objectives to which that action is a means (56).¹⁶ We only need alter a couple details to construct this latter sort of case:

Rat Man II

Although the rat poison at issue is fatal to both rodents and humans, it is actually medically useful for dogs. Furthermore, Rat Man's dog, Robin, is suffering from a painful and, potentially fatal, stomach ulcer, which, his vet has informed him, only this particular brand of rat poison can cure.

From Rat Man's point of view there are now two compelling reasons to buy the poison. He can use it to kill his wife or to cure his dog. If he revises his intention to kill his wife, then, just as in *Rat Man I*, his plan to kill his wife will no longer *be* the reason for which he purchases the poison. But since, in *Rat Man II*, purchasing the poison will still help his dog, our man presumably *will* still purchase the poison in order to help his dog. Thus, he can quite easily bring it about that he purchases the poison not in order to kill his wife, but rather only in order to cure his dog, *by choosing not to kill his wife but to help his dog*. Alternatively, Rat Man can make it the case that he purchases the poison in order to kill his wife, but not to cure his dog, *by choosing to kill his wife but not to help his dog*. In *Rat Man II*, it is in an agent's ability to bring it about that he performs a particular action for one reason rather than another. He does this simply by choosing to perform one or another action.

The reason, then, that *Rescue to Riches* is no counterexample to voluntarism is that I *can* bring it about that I save the man for a reason other than the political situation. I can do this in the exact same way that Rat Man can bring it about that he acts for one reason and not another,

¹⁶ See especially endnote 11 (on 202). When "reasons correspond to two different objectives that the act might achieve," Scanlon writes in this note, "one needs to decide which reason will be the guiding one. But I don't see a similar possibility in cases that do not require a choice between different courses of action." I take it Scanlon thinks that the various reasons for which I might save the man in *Rescue to Riches* don't correspond to different objectives that my action might achieve, but would think that they do in *Rat Man II*.

and, indeed, in the exact same way that I can always bring it about that I perform a given action for one reason and not another: *by altering my choices about which larger courses of action to perform.*¹⁷ These larger actions are less visible in *Rescue to Riches*. Nevertheless, I *can* fix it so that I save the drowning person, but *not* merely to prevent his heir from inheriting, by choosing not merely whether *to* save the drowning man, but also the *conditions* under which I am prepared to save him.¹⁸ I can make choices regarding *when* I am prepared to save him and when I am not. (Moreover, it is certainly no *oddity* that a choice to act may involve undertaking commitments regarding the conditions of one's performance. Think, for instance, of the man who sets out with his lanterns to warn of the Redcoats, planning to hang "One if by land; two if by sea."¹⁹)

If, as stipulated, I would be happy to see the drowning man die, and am able to bring myself to save him *only* in order to prevent his heir from inheriting, then I appear to be committed to following through with the rescue *only* provided certain conditions obtain. As described, I seem committed to *ceasing* my rescue attempt if I find out our drowning victim has lost all his money in the stock market, or has written his presumptive heir out of his will. But if this is right, then I have made a *choice* in pursuing my course of action to rescue the drowning man *only provided that doing so is a necessary means of preventing his heir from inheriting*. And I ought to refrain from such

¹⁷ As I mentioned in footnote 15, given that I cannot *choose* to act-for-a-particular-reason, Scanlon believes that, in order to affect a change in what motivates me, I would have to alter what I see as sufficient reason for acting. However, if I can affect a change in my motivating reasons simply by choosing alternative courses of action, then there is no need for Scanlon to maintain that affecting a change in what motivates me also requires altering what I see as sufficient reason. Moreover, I think it would be strange, by his own lights, if he continued to hold this position, since Scanlon does not think that, in general, I must see sufficient reason in favor of what I do in order to do it. See, for example, Scanlon 1998, 34-35.

¹⁸ Kolodny (2011, 105) makes the suggestion that I can alter my motivating reasons by selecting different conditional intentions. Kolodny leaves largely implicit the point I am emphasizing here: that forming an intention (conditional or otherwise) and, thereby, controlling one's reasons for acting requires not *choosing to act for a particular reason*, but rather *choosing what to do*.

¹⁹ I take this example from Gibbard (2003, 54).

a choice. (Rather, if I ought to make any commitment at all concerning whether to save him should the flow of inheritance change, I should choose to rescue him *regardless of whether doing so is necessary for preventing his heir from inheriting*.) Thus, in acting for the reasons that I do, I *do* make a choice I ought not to: I pursue a plan to rescue a drowning victim only provided that doing so is a necessary means to preventing his heir from inheriting. And it is perfectly up to me *not* to make this choice. Consequently, although my motivating reasons do render me blameworthy, they are also under my control.²⁰

So skepticism about choosing our motivating reasons turns out to be misguided: Agents *can* bring it about via choice that they perform particular actions for some reasons as opposed to others. But the claim that they can't certainly *seems* compelling, which is why cases in which agents' motivating reasons make them blameworthy present a formidable challenge to voluntarism. The claim seems compelling because, although agents *do* bring it about via choice that they perform actions for the particular reasons that they do, they don't do so *by* thinking about the various reasons for which they might perform a given action and then choosing to be motivated by some of those reasons rather than others. Rather, they bring it about that they perform the actions that they do for the particular reasons that they do simply *by choosing what to do*.

²⁰ Although I tried to head off this interpretation in the case's rendering, one may have imagined my psyche slightly differently. Perhaps you imagined that I am committed to saving the drowning man insofar as doing so *is* a means of preventing his heir from inheriting, but just *haven't committed* one way or the other regarding what I will do if rescuing him ceases to be necessary to this end. On this alternative interpretation, my intuitions regarding whether I'm blameworthy start to fail me. However, this does not show that agents can be blameworthy *without* making the wrong choices. Rather, *whether* I am blameworthy seems to intuitively depend on whether I *ought to have* made a commitment concerning what I'd do if the rescue ceased to be a necessary means of preventing inheritance – which in turn depends on details that haven't been specified.

But how exactly do an agent's choices about what to do determine what that person's motivating reasons are? In all of these cases, the connection between the reasons for which our subjects act and the choices they make is very tight. It is not simply that an agent's larger plan "supplies the reason" for which he acts because what his motivating reasons are is a downstream *consequence* of some choice that he has *previously* made. (After all, that Rat Man has previously planned to kill his wife wouldn't necessarily imply that those plans were operative *right now*.) Rather, what makes a difference to which considerations an agent's motivating reasons are, are the (larger) choices he is *presently* making at the time at which he performs an action. When Rat Man purchases the poison, he does so *in order to kill his wife* only if, in purchasing the poison, he is also *continuing* to make the choice to kill his wife. I therefore suggest that agents have voluntary control over what their motivating reasons are because the reasons for which they act are actually *constituted* by choices they are presently making either to pursue larger plans or about the conditions under which they will act.²¹ To be motivated to act by the aim of killing one's wife *is* to be choosing to kill her. Similarly, when I save a drowning man only because of the political situation, my motivating reason is *constituted* by my ongoing commitment to carry out the rescue only in limited circumstances.²²

Can agents choose their motivating reasons then? I suppose it depends what we mean. If we mean, "Can an agent first decide to perform an action, and then decide to regard certain considerations her reason for doing so?" then the answer is "No." But if we simply mean,

²¹ If an agent is committed to acting no matter what the conditions and pursuing no further plans, then he is performing that action unconditionally and for its own sake.

²² Of course, that a person is motivated by some consideration – and thus making a choice – doesn't imply that she will *continue* to make that choice and thus *continue* to be so-motivated. So the view that one's motivating reasons are constituted by one's present choices doesn't imply that people always *accomplish* whatever goals they are choosing to pursue in being motivated by some consideration.

“Can agents control which considerations they act for through their capacity for making choices?” then the answer is “Yes.” It’s just that we control which considerations motivate us not by thinking about particular considerations and attempting to will ourselves to be motivated by *them* rather than others, but rather simply by choosing which plans to pursue at any particular moment.²³

Indeed, to see the fuller implications of this point, we should remember that talking of agents acting on the basis of *particular* considerations they see as reasons is just a shorthand, and not a particularly perspicuous one. When we speak of an agent being motivated by some consideration, it is not as if we mean that the *only* consideration that has had any bearing on what the agent is doing is *that* consideration. Rather, what motivates an agent’s performance of an action done for reasons is not a single consideration, but rather how all the various considerations that might be relevant to what she should do in that moment bear influence with her *on balance*. (What we usually call the “reasons for which” she performs a particular action are just whatever considerations seem salient as *tipping the balance* in favor of her performance of that action.) Thus if you agree with me that which reasons motivate an agent is constituted by her choices, then we should conclude that my choices don’t merely constitute whichever considerations end up weighing heavily enough with me so as to tip the balance in favor of performing some action. Rather, my choices also constitute whatever bearing *any* consideration that could influence how I act has on me in the present moment – including that any number of those considerations bear no weight in what I do at all. What one is

²³ Walen (2006) also recognizes that agents can determine which intentions they form by making choices about which goals to pursue. Walen makes this observation in the course of arguing that it can be impermissible for agents to form certain intentions. He doesn’t discuss its implications for control over our attitudes more generally, or for blameworthiness.

presently choosing to do, the thought is, is what makes determinate the motivational influence held by any consideration in one's practical decision-making. (This includes making determinate one's "motivating reasons" – the considerations that weigh heavily enough in favor of what one is doing so as to tip the scales in favor of doing it.)

I have said that constitutive control is control that we have over some of our attitudes by constituting them via our choices about what to do. We can now be more precise about what this means. Constitutive control is neither quite direct nor indirect. Like our behavior that we control directly, attitudes under constitutive control come about *immediately* through our choices. But like the consequences of our behavior that we control indirectly, attitudes under constitutive control are not themselves the *objects* of our choices; we bring about what we constitutively control by choosing what to *do*. The motivational weight any consideration holds in one's practical reasoning is determined by one's choices to act. But sometimes a person's attitudes just consist in such facts about her. (We have just seen that the ascription of *motivating reasons* to a person consists in such facts. In the subsequent section, I will discuss additional attitudes that are at least sometimes similarly constituted.) When they do, they are attitudes or aspects of her character over which she has constitutive control.

One may object that we *can't* constitute the motivational influence that various considerations have on us by choosing what to do, because particular considerations often bear weight in our practical reasoning *before* we choose to act. For instance, multiple considerations surely carry various weights for and against what to do *during deliberation*. There are two senses in which this is true, but neither undermines the proposal that our choices constitute what motivational weight we give to what: First, most of the time, when one deliberates, one does so under the

guise of a larger course of action that one has *already* chosen to perform. So, for instance, if Rat Man has *already* begun to carry out his plan to murder his wife, then that the poisons and pesticides are in aisle four may *already* bear enough weight with him so as to determine that he shops in that aisle. The initial motivational force of this reason *is* constituted by a choice of Rat Man's, but it is Rat Man's choice to kill his wife – not his choice to shop in aisle four – that constitutes it. That said, it is open to Rat Man, at any moment, to change his mind. That he *continues* to choose uxoricide by simply proceeding to aisle four makes determinate this reason's *continuing* to bear the same influence.

Second, an agent may, prior to making a choice, *see* some consideration as counting with a certain weightiness towards performing a particular action, and yet not ultimately give that consideration the corresponding weight in determining what she *does*. For instance, some consideration may present itself to an agent as *prima facie* bearing a certain normative force, and yet she may choose *not* to give that consideration the weight it initially appears to her to demand, because, upon reflection, she finds her initial impression of its normative import to be mistaken. Additionally agents may sometimes see some consideration as bearing a certain normative weight and yet *never* give it a corresponding amount of motivational influence on what they do, before, but also during and after they act. (In such cases agents act *akratically*.)

So we should distinguish between what an agent gives *normative* weight to and what she gives *motivational* weight to. My claim here is simply that an agent's choices constitute what she gives *motivational* weight to – *not* that her choices constitute what she gives normative weight to, insofar as this diverges from what she allows to motivate her. But I also don't think agents are intuitively blameworthy for what they merely *see* (or fail to see) as reasons, if those reasons

don't also motivate them. Huck Finn, for instance, *isn't* intuitively blameworthy given that he doesn't actually turn Jim in. And, as I'll elaborate in the next section, I don't think what an agent *prima facie* sees as reasons can make her blameworthy independently of any actual determination of her will.²⁴

²⁴ One may worry that agents *can't* give motivational weight to considerations that don't strike them as bearing *any* normative weight, and that this may present a problem for my view. If being motivated by some consideration requires seeing that consideration as bearing at least *some* normative weight, but agents can't control which considerations strike them as bearing any normative weight, then this seems to imply that agents *never* have voluntary control over what they give motivational weight to. But it isn't true that, in general, having voluntary control over something requires having voluntary control over whether one meets the preconditions on which that control depends. In order for Rat Man to have control over whether he kills his wife, it is necessary for him to have been born. But he doesn't need to have had control over whether he was born in order to have control over whether he kills his wife. Thus, that one must be able to see *something* against hurting someone in order to be motivated not to hurt her does not imply that one must have *voluntary control* over whether he sees something against hurting her in order to have control over whether this consideration motivates him. Moreover, I think we can just assume that, insofar as Rat Man is supposed to be eligible for blame at all, so long as it *occurs* to him that he might refrain from killing his wife, then he will see at least *prima facie* reason to do so. Plausibly, seeing at least *prima facie* reason not to hurt any person is part of the basic capacity for moral reasoning that one must possess in order to *ever* be a fitting target of blame. (I discuss what the voluntarist should say about cases where it just *doesn't* occur to agents to do the right thing at the beginning of the subsequent section.)

What if, although Rat Man sees *prima facie* reason to refrain from killing his wife, he just doesn't see sufficient reason to do so? In such a case, even if Rat Man can't *make* himself see sufficient reason to refrain from killing her, he should still be *able* to make the right choice. After all, agents do sometimes seem to act akratically, choosing to do things they don't see sufficient reason to do. Thus, although Rat Man would intuitively be blameworthy for what he does in such a case, what he does will also be under his control.

There are some voluntarists, however, who will argue that agents are *not* blameworthy for acting wrongly if they have false beliefs about what they are obligated to do (at least where this "moral ignorance" does not derive from some earlier culpable failure to adhere to their procedural epistemic obligations) (Zimmerman 1997; Rosen 2003, 2004; Levy 2009). Their thought is that (i) agents can only be held accountable for what it is *reasonable to expect them to do* and (ii) it *isn't* reasonable to expect (non-culpably) morally ignorant agents to meet their obligations, because morally ignorant agents lack the capacity to *rationally* choose to do what they ought to. (For this line of argument see, for example, Rosen 2004, 306, and Levy 2009, 735-739.) But other voluntarists *deny* (ii) and hold that it *is* reasonable to expect morally ignorant agents to act rightly because agents *can* rationally choose to do things they don't now think they should. (See, for example, Wallace 1994, especially 163-164, and Kane 1996, especially 132-133.) So, as far as *voluntarism* commits one, intuitions that agents like Rat Man can be blameworthy for acting wrongly even when they have false beliefs about what is required of them can be perfectly correct.

Interestingly though, I think that probably the best *reason* for a voluntarist to reject (ii) actually involves appealing to the thought that an agent's choices can constitute not just what motivates her, but also what she *sees* herself as having sufficient reason to do. (See Kane 1996, 2007.) (ii) says that agents can't rationally do what they don't see sufficient reason to do. But this seems to presuppose that what an agent sees all-things-considered reason to do is always settled prior to her choosing what to do. And this assumption seems false. Sometimes it is only *in* making a choice that a person settles what she thinks she should do. In such cases an agent's choices *do* appear to constitute what she sees sufficient reason to do. So I actually think the idea that an agent's choices sometimes constitute what she gives *normative* weight to merits further discussion in the context of disagreements *between* voluntarists about culpability in cases of moral ignorance.

Our discussions of *Rescue to Riches* and *Rat Man II* both employ what we can now regard as a general strategy for replying to any cases in which an agent's motivating reasons make her blameworthy. An agent's reasons for acting always indicate what she is committing herself to bringing about in acting: what she is choosing to do under what conditions, or what *plans* she choosing to pursue. (Or, some would say, her *maxim*.) Thus, in any case where an agent is acting for objectionable reasons, those reasons will indicate that she is either pursuing plans she ought not to be, or failing to pursue plans that she should. But whether she makes or fails to make such choices is under her voluntary control.

IV. Blameworthiness for Mental States

But perhaps "right thing for the wrong reason cases" are not really the strongest weapons in the non-voluntarist arsenal to begin with. One might think the cases most threatening to voluntarism aren't those in which agents appear to be blameworthy for performing actions for particular reasons. Rather, the cases about which I should be most concerned are those in which individuals appear blameworthy *just* for their mental states (or lack thereof) in isolation.

The voluntarist can reply to some of these cases using tools already at her disposal. For example, sometimes it doesn't occur to an agent to do what she ought to do at a particular moment, or agents forget things that are important to other people, like their birthdays or anniversaries. Moreover, intuitively, persons can sometimes be blameworthy in such cases, not merely for any inaction due to such lapses but also for the mental lapses themselves. Yet, some non-voluntarists have argued, since agents don't *voluntarily* forget or remember, voluntarists cannot account for what makes such people blameworthy.²⁵

²⁵ See, for example, Smith (2005).

I don't think that the notion of constitutive control that this paper is about particularly aids the voluntarist in responding to this type of purported counterexample. But I also think that voluntarists have responded adequately to this type of case already. The gist of the reply I favor is that agents *do* have some voluntary control over what they remember, because they are able to make choices to help remind themselves of things.²⁶ Agents thus sometimes have obligations to remind themselves of things, and can be blameworthy for forgetting when their forgetting is traceable back to a failure to take steps to remind themselves when they could and should have. If an agent *has* taken all the steps that she was obliged to take, and yet simply still does not remember, then she is *not*, in fact, blameworthy for forgetting.²⁷

Some additional cases involving persons purportedly blameworthy for their mental states are ones about which I think the voluntarist and non-voluntarist may simply have divergent intuitions. For example, I mentioned earlier the case of a man who believes that women naturally have less aptitude for math. A voluntarist will hold that such a person is blameworthy

²⁶ See, for example, Fischer and Tognazzini (2012) and Levy (2005).

²⁷ One may object that if it *never occurs* to an agent to take steps to remind herself of something, then she cannot have had control over her failure to take such steps, and thus cannot be blameworthy for forgetting. But some voluntarists will hold that agents *do* have voluntary control in such cases so long as their failures to remind themselves of what to do are constituted by their making *other* voluntary choices instead (Fischer and Tognazzini 2012). Alternatively (on my favored reply), the voluntarist can just respond to this objection by *conceding* that if it *never* occurs to an agent to take steps to remind herself of something, and *this* failure isn't itself traceable back to any point at which it *did* occur to the agent to do something, the probable result of which would have been that she remembered (or remembered to do something else which probably would have led to her remembering) then the agent *isn't* blameworthy. But then, cases in which agents are intuitively *blameworthy* for forgetting something are not generally like this; normal agents *do* have the opportunity to remind themselves of the things they are supposed to remember, because it generally *does* occur to them that, say, their friends' birthdays are approaching, or (if not that then) that their friends are people to whom they ought to demonstrate care. Moments when such things occur to us present opportunities take action that will help us to remember to demonstrate adequate care to the people in our lives. If there is a person to whom it *never* occurs that others require demonstrations of care, then I take it that said person is either not a morally competent agent (and so doesn't meet the preconditions for blameworthiness), or has had some sort of rare and abnormal brain glitch (in which cases we *shouldn't* hold her responsible for forgetting).

for his belief only if its acquisition or maintenance is the (foreseeable) result of some decision or preventable failure to attend to evidence that ought to have been attended to. But some non-voluntarists may wish to hold that such a person is blameworthy regardless of whether such a condition is satisfied. If this is so, I think the voluntarist should simply stand her ground. After all, part of the *point* of being a voluntarist about blameworthiness is to recognize that there are many instances in which, although others may find our attitudes disappointing or hurtful, we are nevertheless not blameworthy for them, precisely because we cannot voluntarily control them.

There are still other cases, however, in which it is less plausible for the voluntarist to claim that agents are simply *not* blameworthy in spite of their problematic mental states. It is here that I think the notion of constitutive control can help the voluntarist once again. Two varieties of such case support non-voluntarism most strongly: cases in which people appear blameworthy for failing to care adequately about others, and cases where people appear blameworthy for holding objectionable attitudes towards others' suffering.²⁸ Consider an example of the first sort:

Sister's Goals

Growing up, my younger sister was very devoted to her traveling soccer team. I, on the other hand, never liked soccer, and never took much of an interest in my sister's athletic achievements or endeavors. She now resents me for my failure to care about her goals.²⁹

A voluntarist could treat this as a case of *indirect* blameworthiness. My failure to care about my sister's goals may be blameworthy, she could say, only when and because it is the consequence

²⁸ Adams (1985) and Graham (2014) both raise these sorts of cases.

²⁹ Pun intended.

of some earlier failure of choice. Generally, sisters have obligations to foster care for one another, and so a failure to care can indicate a(n impermissible) failure to take the steps necessary to foster that care. But I take it that *Sister's Goals* presents a challenge to voluntarism in part because we can imagine my sister rejecting this interpretation of her own resentment, and insisting that I am blameworthy simply because I don't care about supporting her goals. It doesn't *matter* whether any earlier choices or omissions of mine may have contributed to fostering my uncaring attitudes, we can imagine her thinking. The problem for which she blames me is simply that I don't care. If it is legitimate for my sister to blame me on this count, then one might think the case presents a counterexample to voluntarism, because it is not up to me whether I care. I cannot *choose* whether or not I value my sister's athletic achievements and so it cannot be wrongdoing on my part if I fail to.

But in replying earlier to the "right thing for the wrong reason" cases, we revealed that even if we cannot *choose whether to have* certain attitudes, taking a managerial stance towards our attitudes is not always our only way of controlling them. We can also control at least some of our attitudes by making choices about what to do that constitute those attitudes. Caring is one such attitude. For whether I care often *is* just a matter of what I prioritize via my choices. Perhaps, when growing up, I only ever gave my sister rides to practice when our mother forced me to. Perhaps I often got together with my friends in the afternoons, and never once planned to do *conditional* on my sister not having a game that day. These perfectly voluntary choices may together constitute my failure to care about my sister's soccer-related goals. Thus, in *Sister's Goals*, I may be blameworthy for failing to care adequately about the things that are important to my sister, not because I have impermissibly omitted *to choose to care*, but rather

because my failure to have the requisite attitude is constituted by a series of failures to make the right choices about what to do.

This is not to claim that *every* time we speak of what an agent “cares” about, we are referring to attitudes under her constitutive control. Think of a mother suffering from postpartum depression, who nevertheless sees to it that her child is properly looked after and who assiduously goes to therapy to work on her disease. There is a sense in which such a mother does not care for her child. But she is not blameworthy for her failure to care in this sense. (People may *blame* her, of course, but she doesn’t *merit* such censure; wanting her to feel bad for her condition would, intuitively, be unfitting.) For there is another sense in which she clearly does care, which is constituted by her choice to take steps towards treatment. Moreover, it is whether she cares in this latter sense – the sense that *is* under her constitutive control – that makes a difference to whether she’s blameworthy.

The other cases that seem to strongly favor non-voluntarism are those in which agents appear blameworthy for having objectionable attitudes concerning others’ pain: particularly cases in which agents hope, wish, or desire for others to suffer harm or pain, or take pleasure in their doing so. Non-voluntarists claim that we can be blameworthy for having such attitudes, although it is not up to us whether we do. Consider:

Torture

*Carmen watches as Maria is led off to be tortured – an event that it is not in Carmen’s power to stop. As Carmen observes Maria, she finds herself hoping that the torture will be especially painful. When she hears later that Maria suffered a great deal, Carmen is pleased.*³⁰

³⁰ Here I combine two cases proposed by Graham (2014).

We may suggest on the voluntarists' behalf that the attitudes that make Carmen blameworthy in this case resulted (foreseeably) from her poor choices, and were thus under her *indirect* control. For all we've been told, Carmen may have chosen to enroll in a school for bounty hunters and assassins, and thereby come to develop an aesthetic appreciation for various sorts of suffering. She is blameworthy for her dubious attitudes, we could then contend, only because they resulted (predictably) from her equally dubious choices. But (once again) the non-voluntarist will rightly protest that appeal to indirect control is inadequate for explaining this case away. After all, in *Torture*, it seems reasonable to blame Carmen, *regardless* of whether her attitudes resulted foreseeably from any of her actions. Perhaps she just came to despise Maria and wish her ill through a series of interactions with her, although there was no point at which it was predictable that she'd develop such rancor. Even so, Carmen's attitudes seem blameworthy.

But it is important to point out that there are two senses in which one might be said to "desire" some object, and, correspondingly two senses in which one might take "pleasure" in it. There is desire in the sense of finding oneself inclined toward something, or being struck with yearning for it: of *prima facie* seeing reason to bring it about. Such desire is passive, affective, and seemingly out of direct control – as is the sort of pleasure that is derived from its satisfaction, the prospect of which is its ground. Indeed, non-voluntarists *must* be thinking of this sense of desire when they contend that although our desires can make us blameworthy "they are not cases of trying, choosing, or meaning, as is illustrated by the fact that if I simply desire to do something, it remains a question whether I will try or choose or mean to do it" (Adams 1985, 9-10). For there is another sense of "desire" in which to desire that some state of affairs come about *just is* to choose it as one's goal or aim. This is the sense of "desire" we

invoke when we say that the sweet-tooth's choice to refrain from eating proved that she *didn't* really desire the candy bar after all.³¹ When I desire in *this* sense, no question *does* remain about whether I will choose what I desire, since what I desire in this sense is *constituted* by my choice to pursue some aim or goal. And whether I experience the pleasure that consists in *this* desire's satisfaction is up to me too. For it only results if I form such desire, and whether I do so *is* under my constitutive control.

Torture does not specify the sense in which Carmen desires Maria's pain, or the sense in which she takes pleasure in it. But when we return to the case these distinctions in hand, we find we care about Carmen's desires in the sense of her goals, not in the sense of her inclinations. If Carmen is committed to bringing Maria's suffering about, or takes pleasure in it as a moment of all her plans coming to fruition, then she certainly seems blameworthy for such mental states. If she *merely* feels a flash of pleasure at the thought of Maria's pain, however, then she doesn't necessarily. To be sure, if Carmen feels a twinge of delight at Maria's suffering with no *corresponding* desire to be *rid* of this revolting impulse, then we may certainly feel that she deserves our sanction. But when we stipulate that Carmen's desires and pleasures are ones that she herself longs to be rid of, and has done everything in her power to escape, then the intuition of blameworthiness dissipates. If we imagine her a sadist that has second-order desires, above all else, *not to be*, then I don't think we can continue to see as condemnable – although we may certainly regard her situation as unfortunate, and her first-order passions as distasteful. The idea of wishing that she suffer for her first-order feelings, however, just seems unmerited and overblown.³²

³¹ For the distinction between these types of desire, see, for example, Nagel (1970, 29).

³² Rosen (2004, 302-303) shares the intuition that when a person suffers from *schadenfreude* that he disavows, then "it's simply obtuse to hold him responsible for his reaction."

This, then, is what I think is going on: Although Carmen may be blameworthy for her desires in the sense of yearnings for the goals she sets herself, and for the pleasures of attaining such goals, whether she possesses attitudes of *these* sorts is under her constitutive control. On the other hand, Carmen cannot be directly blameworthy for her desires or pleasures that are simply passive strikes of fancy – *although she may very well be blameworthy for failing to desire to be rid them*. But then, whether she possesses this second-order desire is under her voluntary control too! *This* is a desire that Carmen can constitute through her choices. She can either passively notice her baser instincts or she can choose to set herself against them. Thus, when we clarify which of Carmen’s attitudes are under her constitutive control and which are not, whether she is blameworthy seems to depend on whether she possesses attitudes over which she *does* have constitutive power.

The best cases for non-voluntarists are those in which agents appear blameworthy for either possessing or lacking attitudes: particularly certain other-regarding desires or cares. But terms like “caring” and “desiring” (and “being committed” and “valuing”) are loose. Sometimes all it is to “care about” or “be committed to” some thing, person, or goal is to choose to conduct oneself respectfully towards that thing, person, or goal. Sometimes all it is to “desire” or “value” some end is to choose to pursue it. And sometimes all it is to fail in these respects is to omit to so-choose. If you think that it is not in your power to value some person, or to care about her pain, just imagine a person – any real person – and ask yourself: Can I commit myself to taking that person’s rights seriously, and to only making plans that don’t diminish her well-being? When one reflects like this, I submit, one finds that one *can* make such

commitments if one chooses to. Such choices *constitute* some of one's attitudes towards that person.

Thus, in considering putative counterexamples to voluntarism, we must remember that some of our attitudes can be constituted by our choices to act. This is of course not true of *all* of our attitudes. (And I certainly don't mean to imply that the terms I mentioned in the paragraph above are used *exclusively* to refer to attitudes that are so-constituted.) Taking constitutive control seriously, however, should ultimately lead us to favor voluntarism. This is because, when we disambiguate the cases where agents *do* appear blameworthy for their attitudes, blameworthiness appears to depend on precisely those attitudes that agents *can* control – if not indirectly, then constitutively.

V. Conclusion

The lesson here is that, although it is convenient to cast the debate between voluntarists and non-voluntarists as one regarding whether “responsibility is ultimately for *being*, rather than for doing” (Hieronymi 2014, 5), technically this dichotomy is false. Some ways of being *are* what we do. The voluntarist claim is not that individuals can never be blameworthy for aspects of who they are. Rather, the claim is that individuals can only be blameworthy for those aspects of who they are that *are* under their voluntary control. Thus the non-voluntarists are right that we can be blameworthy in virtue of our characters, and particularly for being problematically oriented towards other people. But there is no need for the voluntarist to disavow responsibility for these parts of ourselves. This is because, when we take constitutive control

seriously, the most condemnable aspects of our characters turn out to be perfectly within our control.³³

³³ I am grateful to two anonymous reviewers for their rich and careful commentary. For helpful conversations and notes on earlier drafts, thanks to Olivia Bailey, Jeff Behrends, Sanford Diehl, Sam Dishaw, Lidal Dror, Nick French, Ned Hall, Elizabeth Harman, Douglas Kremm, Emma McClure, Lowry Pressly, Mathias Risse, Alison Simmons, Angela Sun, and Kate Vredenburg, as well as to participants in Athena in Action: a Networking and Mentoring Workshop for Graduate Student Women in Philosophy, the Edmond J. Safra Center for Ethics Graduate Seminar, and the Harvard Moral and Political Philosophy Dissertation Workshop. Especial thanks to Selim Berker, Christine Korsgaard, and Tim Scanlon for multiple rounds of written commentary and illuminating discussion.

References

- Adams, R. (1985). Involuntary sins. *Philosophical Review* 94(1), 3-31.
- Audi, R. (1991). Responsible action and virtuous character. *Ethics* 101(2), 304-321.
- Carlsson, A. B. (2017). Blameworthiness as deserved guilt. *The Journal of Ethics* 21(1), 89-115.
- D'Arms, J. and Jacobson, D. (2000a) The moralistic fallacy: On the 'appropriateness' of the emotions. *Philosophy and Phenomenological Research* 61(1), 65-90.
- D'Arms, J. and Jacobson, D. (2000b). Sentiment and value. *Ethics* 110(4), 722-748.
- Fischer, J.M. and Ravizza, M. (1998) *Responsibility and control: A theory of moral responsibility*. Cambridge: Cambridge University Press.
- Fischer, J.M. and Tognazzini, N. (2012). The triumph of tracing. In J.M. Fischer *Deep control: An essay on free will and value* (pp. 206-234). Oxford: Oxford University Press.
- Gibbard, A. (2003). *Thinking how to live*. Cambridge, MA: Harvard University Press.
- Graham, P. (2014) A sketch of a moral theory of blameworthiness. *Philosophy and Phenomenological Research* 88(2), 388-409.
- Hanser, M. (2005) Permissibility and practical inference. *Ethics* 115(3), 443-470.
- Hume, D. *Treatise of human nature*. (1740/1978). L.A. Selby-Bigge (Eds.) 2nd Edition. Oxford: Oxford University Press.
- Hieronymi, P. (2008) Responsibility for believing. *Synthese* 161(3), 357-373.
- Hieronymi, P. (2014). Reflection and responsibility. *Philosophy and Public Affairs* 42(1), 3-41.
- Howard, C. (2018). Fittingness. *Philosophy Compass* 13(11), e12542.
- Kane, R. (1996). *The significance of free will*. Oxford: Oxford University Press.
- Kane, R. (2007). Libertarianism. In J.M. Fischer et al. *Four views on free will*. Malden, MA: Blackwell Publishing, 5-43.
- Kolodny, N. (2011). Scanlon's investigation: The relevance of intent to permissibility. *Analytic Philosophy* 52(2), 100-123.
- Levy, N. (2005). The good, the bad, and the blameworthy. *Journal of Ethics and Social Philosophy* 1(2), 1-14.
- Levy, N. (2009). Culpable ignorance and moral responsibility: A reply to FitzPatrick. *Ethics* 119(4), 729-741.

- Macnamara, C. (2015). Blame, communication, and morally responsible agency. In R. Clarke, M. McKenna, and A. Smith (Eds.) *The nature of moral responsibility: New essays* (65-88). Oxford: Oxford University Press.
- McHugh, C. (2017). Attitudinal control. *Synthese* 194(8), 2745-2762.
- McKenna, M. (2013). Directed blame and conversation. In D.J. Coates and N. Tognazzini (Eds.) *Blame: Its nature and norms* (pp. 119-140). Oxford: Oxford University Press.
- Moody-Adams, M. On the old saw that character is destiny. (1990) In O. Flanagan and A. Rorty (Eds.) *Identity, character, and morality* (pp. 111-132). Cambridge, MA: MIT Press.
- Nagel, T. (1970). *The possibility of altruism*. Princeton: Princeton University Press.
- Pereboom, D. (2014). *Free will, agency, and meaning in life*. Oxford: Oxford University Press.
- Rosen, G. (2003). Culpability and ignorance. *Proceedings of the Aristotelian Society, New Series* 103(1), 61-84.
- Rosen, G. (2004). Skepticism about moral responsibility. *Philosophical Perspectives* 18(1), 295-313.
- Rosen, G. (2015). The alethic conception of moral responsibility. In R. Clarke, M. McKenna, and A. Smith (Eds.) *The nature of moral responsibility: New essays* (65-88). Oxford: Oxford University Press.
- Scanlon, T.M. (1998). *What we owe to each other*. Cambridge, MA: Harvard University Press.
- Scanlon, T.M. (2008). *Moral dimensions: Permissibility, meaning, blame*. Cambridge, MA: Harvard University Press.
- Scanlon, T.M. (2013) Interpreting blame. In D.J. Coates and N. Tognazzini (Eds.) *Blame: Its nature and norms* (pp. 84-99), Oxford: Oxford University Press.
- Sher, G. (2006) *In praise of blame*. Oxford: Oxford University Press.
- Shoemaker, D. (2003). Caring, identification, and agency. *Ethics* 114(1), 88-118.
- Shoemaker, D. (2011). Attributability, answerability, and accountability: Toward a wider theory of blameworthiness. *Ethics* 121(3), 602-632.
- Smith, A. (2005). Responsibility for attitudes: Activity and passivity in mental life. *Ethics* 115(2), 236-271.
- Smith, A. (2008). Control, responsibility, and moral Assessment. *Philosophical Studies* 138(3), 367-392.

- Smith, A. (2012). Attributability, answerability, and accountability: In defense of a unified account. *Ethics* 122(3), 367-392.
- Thomson, J. (1991) Self-defense. *Philosophy and Public Affairs* 20(4), 283-310.
- Wallace, R.J. (1994). *Responsibility and the moral sentiments*. Cambridge, MA: Harvard University Press.
- Walen, A. (2006). The doctrine of illicit intentions. *Philosophy and Public Affairs* 34(1), 39-67.
- Watson, G. (1996). Two faces of responsibility. *Philosophical Topics* 24(2), 227-248.
- Wolf, S. (2011). Blame, Italian style. In R.J Wallace, R. Kumar, and S. Freeman (Eds.) *Reasons and recognition: Essays on the philosophy of T.M. Scanlon* (pp. 332-487). Oxford: Oxford University Press.
- Zimmerman, M. (1997). Moral responsibility and ignorance. *Ethics* 107(3), 410-426.