# Control, Attitudes, and Accountability*

**Douglas W. Portmore**

**Abstract:** It seems that we can be directly accountable for our reasons-responsive attitudes—e.g., our beliefs, desires, and intentions. Yet, we rarely, if ever, have volitional control over such attitudes, volitional control being the sort of control that we exert over our intentional actions. This presents a trilemma: (Horn 1) deny that we can be directly accountable for our reasons-responsive attitudes, (Horn 2) deny that φ's being under our control is necessary for our being directly accountable for φ-ing, or (Horn 3) deny that the relevant sort of control is volitional control. This paper argues that we should take Horn 3.

Volitional control is the sort of control that we exert directly over our intentional actions and indirectly over those things that we manipulate via such actions. A subject who has direct volitional control over whether she φs: (1) can φ at will and, thus, can φ simply by trying, intending, or otherwise willing to φ; (2) can φ for whatever reason she takes to be sufficient for φ-ing and, thus, can φ to win a bet or to please her partner; and (3) can—at least, typically—choose when to φ and, thus, can choose whether to φ now or later.

Interestingly, we don't typically, if ever, exert direct volitional control over our reasons-responsive attitudes—e.g., our beliefs, desires, and intentions.[1] For the sort of control that I exert over whether, say, I believe that Aristotle went for a swim on his 30th birthday is unlike the volitional control that I exert over whether I intentionally touch my nose (McHugh 2017). I can do the latter but not the former at will. I can do the latter but not the former to win a bet. And whereas I can choose when to touch my nose, I cannot choose when (or even whether) to believe that Aristotle went for a swim on his 30th birthday. But even though our reason-responsive attitudes are not typically, if ever, under our volitional control, we can, it seems, be responsible for them. It seems, for instance, that I can be responsible for wanting to hurt someone, for accepting a proposition on insufficient evidence, and for intending to do what I take to be incompatible with my ends. And it's not just that these attitudes can rightly be

* Promised to the fifth volume of *Oxford Studies in Agency and Responsibility*.

[1] Reasons-responsive attitudes include all and only those mental states that a rational subject will tend to have, or not to have, in response to (apparent) reasons for or against having them. Thus, belief, but not hunger, is a reasons-responsive attitude.

attributed to me. What's more, I can rightly be held to account for having formed them. Indeed, it seems that I can be *directly* accountable for such attitudes—accountable for them without this being in virtue of my being accountable for something else. Yet, this may seem problematic given that accountability, unlike mere attributability, entails liability to reward or sanction (see WATSON 1996 and SHOEMAKER 2015, 87). For no one deserves reward or sanction in virtue of something that wasn't, in the relevant sense, under her control. So, if, as it seems, I can be directly accountable for my reasons-responsive attitudes, the relevant sort of control must not be volitional control.

We face, then, the following trilemma: (Horn 1) deny that we can be directly accountable for our reasons-responsive attitudes, (Horn 2) deny that an event's being under our control is necessary for our being directly accountable for it, or (Horn 3) deny that the relevant sort of control is volitional control. I'll argue that we should take hold of the third horn. But, first, I'll explain why we should be loath to impale ourselves on either of the first two horns.

**1. The First Horn: Deny Direct Accountability for Our Reasons-Responsive Attitudes**

Admittedly, volitionalists—those who think that volitional control is necessary for accountability—can explain our accountability for events that weren't under our direct volitional control by tracing their causes back to events that were. For instance, they can explain why I'm accountable for forgetting my daughter's birthday by tracing its cause back to my intentionally choosing not to put the relevant reminder in my calendar. So, volitionalists might hope to employ the same strategy in explaining our intuition that we can be accountable for our reasons-responsive attitudes. There are, however, three problems with this: (1) it won't always work, (2) even when it does work it forces us to accept a counterintuitive view about what we're ultimately accountable for, and (3) it's unmotivated in the case of reasons-responsive attitudes given that such attitudes are themselves responsive to reasons. I'll take each in turn.

To see that the tracing strategy won't always work, consider *Earth's Age:* a well-educated woman of normal intellect named Chen believes, as she was brought up to believe, that the Earth is no more than a few thousand years old. But, being well-educated, Chen is well aware of all the scientific evidence to the contrary. What's more, she possesses the relevant rational

capacities for recognizing and responding appropriately to the reasons that such evidence provides. Given this, she seems accountable for not having responded appropriately to her reasons—that is, for not having abandoned her ill-founded belief. Yet, she couldn't have abandoned her belief at will; for abandoning it in response to the decisive reasons that she had for doing so was not something over which she exerted volitional control. Consequently, the volitionalist must trace the source of her accountability back to something over which she did exert volitional control. That is, the volitionalist must trace it back to her having failed to perform some intentional act that would have caused her to abandon that belief (A. M. SMITH 2015).

But this strategy won't work if we assume, as I will, that the only intentional act that would have caused her to abandon the belief was one that she was required to refrain from performing. Assume, then, that given her penchant for conspiracy theories, the only intentional act that would have caused her to abandon this belief was her reading a book by some quack claiming that the Bible was written and propagated by the CIA for the purposes of controlling the masses. And assume that she's required to refrain from reading such books both because they contain a lot of dangerous misinformation that she's liable to believe and because she promised her mother that she would not read such books. Here, then, is a case where Chen should have responded appropriately to her reasons and abandoned her belief, and yet the volitionalist can't even claim that there was something that she ought to have done intentionally to have caused this. For, in this case, the only intentional act that would have caused this was one that she was required to refrain from performing. So, the tracing strategy fails in this case.

Of course, the tracing strategy will work in other cases, such as *Earth's Age 2*—a case that's exactly like the original except that, in this case, Chen could have caused herself to abandon her belief by permissibly and intentionally choosing to surround herself by clear-thinking friends who would have persuaded her to abandon her belief. The problem, though, is that this strategy forces us to accept an implausible view about what she's ultimately accountable for. According to this strategy, what Chen is ultimately accountable for is her failure to perform the acts that would have resulted in her being surrounded with clear-thinking friends. She would not, then, be accountable for her failure to respond appropriately to

her evidence except insofar as her failure to perform these acts led to this failure. Thus, the norm that she would ultimately be accountable for violating is not the epistemic norm requiring her to believe in accordance with her evidence but the practical norm requiring her to act in accordance with her practical reasons. But, intuitively, it seems that what she's ultimately accountable for is violating an epistemic norm. And this is why, when we interact with people like Chen, we exhort them to respond to their epistemic reasons for abandoning their ill-founded beliefs, not to their practical reasons for doing what would cause them to abandon these beliefs. That is, we appeal to their epistemic reasons, not to their practical reasons.

These are not the only problems with the volitionist's attempt to trace our accountability for our reasons-responsive attitudes back to something over which we exerted volitional control. What's, perhaps, most problematic about this strategy is that it's unmotivated. The motivation for our employing a tracing strategy always lies with our having the intuition that a subject is accountable for something that wasn't itself responsive to reasons. Thus, it's only because the something in question isn't itself responsive to reasons that we feel the need to trace the source of the subject's accountability back to something that was. To illustrate, consider *The Drunk Driver:* a drunk driver inadvertently hits a pedestrian due to her blurred vision and impaired motor skills. Here, there's a motive for tracing, because it seems that she can't be directly accountable for failing to apply the brakes sooner given that she was incapacitated and, consequently, incapable of recognizing and/or responding appropriately to the reasons for doing so. Consequently, we feel the need to trace back the source of her accountability to something that was responsive to her reasons—something such as her decision to start drinking without first arranging for a designated driver.[2] And, here, it does make sense to insist that she should have recognized and responded appropriately to the reasons that she had for doing so (A. M. SMITH 2015), because, unlike in the case of the reasons for braking sooner, she was capable of recognizing and responding appropriately to these reasons.

[2] Note that it wouldn't be sufficient to trace back the source of her accountability merely to something that was under her volitional control. For we wouldn't think her accountable for having hit the pedestrian in virtue of her having intentionally chosen to start drinking without having first arranged for a designated driver unless she was capable of recognizing and responding appropriately to the reasons she had for doing so.

Note, then, that the motivation for tracing in *The Drunk Driver* is completely lacking in any case in which the event in question is itself responsive to reasons. In any such case, there's no need to trace the source of the subject's accountability back to some event that was responsive to reasons. For, in any such case, the event in question is itself responsive to reasons. So, the volitionalist's strategy of tracing the source of our accountability for our reasons-responsive attitudes to some event over which we exerted direct volitional control is unmotivated.

Of course, I've been assuming that we are, as it seems, often accountable for our reasons-responsive attitudes and that, consequently, it's a problem if the volitionalist can't provide a plausible account of this. Yet someone may be willing to embrace the first horn of our trilemma and simply deny that we are ever (either directly or indirectly) accountable for such attitudes. And, for such a person, the failure of the tracing strategy with respect to reasons-responsive attitudes will cut no ice. But, given that we do seem to be accountable for such attitudes, such a revisionary position should, I believe, be a last resort. And, so, I'll wait until we've considered every alternative before discussing this unpalatable possibility in the concluding section.

**2. The Second Horn: Deny the Necessity of Control for Accountability**

Some may think that we should just respond to our trilemma by denying the necessity of control for accountability. But this, I believe, is implausible. To see why, note that to be accountable for φ-ing is for it to be appropriate to be held to account for it and, thus, to be liable to reward or sanction for it.[3] The reward or sanction needn't come from the law, society, or common opinion, but it must at least come from the approval or disapproval of one's own conscience—see MILL 1991, chap. V. Thus, a subject is accountable for having φ-ed if and only if it would be appropriate for her to feel either prideful or guilty for having φ-ed, depending on whether φ's valence is positive or negative.[4] And this, I'll stipulate, is just what I mean by the

[3] I've purposely used the word 'accountable' as opposed to 'responsible', as some philosophers think that there are several distinct types of responsibility. For instance, David Shoemaker (2015) argues that there are three distinct types: attributability, answerability, and accountability. See also WATSON 1996, who argues for two of these three.

[4] It would be appropriate for her to feel prideful for having φ-ed if and only if her φ-ing was something she ought to have done, and it would be appropriate for her to feel guilty for having φ-ed if and only if her φ-ing was something that she was obligated to have refrained from doing. Also, note that, as I've defined 'accountability', no one can be accountable for a neutral event since it's

phrase 'accountable for having φ-ed'. Nevertheless, I need to explain what I mean by the terms 'prideful', 'guilty', and 'appropriate'.

Like other emotions, pride and guilt have cognitive components in addition to their affective and motivational components. These cognitive components set their appropriateness conditions, such that a given emotion is appropriate if and only if the thoughts that it implicates are fitting or correct.[5] Thus, the relevant sense of appropriateness is not some domain-specific sense, such as that of being morally or legally appropriate. Rather, an attitude is appropriate in the relevant sense just in case the thoughts that it implicates are true. This is what Gideon Rosen (2015) calls the *alethic sense* of 'appropriate'.[6] In this sense, the fear of *X* is appropriate if and only if *X* is dangerous, for the fear of *X* implicates the thought that *X* is dangerous.[7]

But what thoughts are constitutive of feelings of pride and guilt? I take them to be, respectively, the thought that one deserves to experience the pleasantness of feeling prideful and the thought that one deserves to experience the unpleasantness of feeling guilty. To illustrate, take guilt. Its affect is unpleasant, for if the feeling that one's experiencing isn't unpleasant, it can't be guilt.[8] But guilt isn't just any unpleasant feeling associated with certain motivational tendencies. To feel guilt, one must additionally have the thought that one deserves

---

never appropriate for one to feel prideful or guilty for having done something neutral. I do this solely for the sake of simplifying the exposition. If, instead, we want to allow that people can be accountable for neutral events, we can just tweak the right-hand side of my bi-conditional as follows: "it is the sort of thing that it would be appropriate to feel either prideful or guilty about provided it had either a positive or negative valence."

[5] Perhaps, this is too quick. For it doesn't seem appropriate for a worm to fear a bird unless the worm is capable of forming the thought that this attitude implicates—that is, the thought that the bird is a danger to it. So, I should probably qualify the above as follows: it's appropriate for a subject to form an attitude only if she has the option of forming this attitude as well as the thoughts implicated by this attitude. And, given my view of accountability above, this would mean that only those subjects with the capacity for feelings of pride and guilt can be accountable for their actions.

[6] Thus, I'm not saying that pride and guilt are inappropriate whenever it would be immoral to have such attitudes. That would be to commit what Justin D'Arms and Dan Jacobson (2000) call the *moralistic fallacy*. After all, a reaction can be fitting even if morally inappropriate. For instance, laughter can be the fitting response to a funny joke even if that response would be morally inappropriate given that it's a cruel joke.

[7] The idea that it would be appropriate/inappropriate (that is, fitting/unfitting) for a subject to φ is distinct from the idea that it would be fortunate/unfortunate that she φs. Consequently, we must allow that it could be appropriate, say, to fear an animal even if this would be unfortunate given that the animal would then sense this fear and become even more dangerous. Despite its being unfortunate to have this fear, it would, nevertheless, be appropriate so long as it correctly represents its object as dangerous. In general, attitudes represent their objects as being a certain way and are, therefore, appropriate (that is, fitting and correct) to the extent that their representations are accurate. By contrast, an attitude is *fortunate* if and only if good consequences would result from one's having that attitude. For more on this distinction, see CHAPPELL 2012.

[8] I'm not alone in thinking that feeling guilt is essentially an unpleasant experience. See, for instance, CARLSSON 2017 (91), CLARKE 2016 (122), MORRIS 1976 (101), ROSEN 2015 (67, n. 6), and WOLF 2011.

to feel bad given one's failure to live up to some legitimate demand.[9] Thus, a woman with Tourette's may feel bad for having involuntarily uttered some obscenity, but this feeling won't amount to guilt unless she experiences its unpleasantness as at least partially deserved. And, if she had no control over whether she was to utter this obscenity, she shouldn't experience this unpleasantness as even partially deserved. She should instead experience it as she does the unpleasantness of a headache and, thus, as something she should wish to be rid of.

Of course, none of this is to suggest that it's impossible to feel guilty unless one *believes* that one deserves to experience the unpleasantness of guilt. Clearly, one can feel guilty without believing that one deserves to feel this way just as one can fear something without believing that it's dangerous. Although fearing something implicates the *thought* that it's dangerous, one needn't *believe* that it's dangerous in order to have this thought.[10] Having the thought that it's dangerous necessitates only experiencing it as dangerous. And to experience it as dangerous is just for it to strike one as dangerous in the same way that the lines in a Müller-Lyer illusion can strike one as unequal even if one believes that they're equal.[11] Likewise, to have the thought that one deserves to experience the unpleasantness of guilt given that one has failed to live up to some legitimate demand, it's sufficient that this experience strike one as being (at least, partially) deserved.[12] And this is why we experience the unpleasantness of guilt differently than we do the unpleasantness of, say, a headache. Unlike the unpleasantness of a headache, the unpleasantness of guilt strikes us as something that it would, in some respect, be problematic to be rid of. Thus, taking a pill to alleviate one's appropriate guilt seems problematic in a way that

[9] Here, I concur with Darwall and Mill: "Mill calls guilt a kind of 'internal sanction', but it is important to appreciate that guilt is not merely painful, or the (painful) fear of further (external) sanctions (MILL 1998: Ch. III). It is the painful sense of having done wrong, having violated a legitimate demand that comes, not just from someone else, say God, but also that one implicitly makes of oneself, through blaming oneself in feeling guilt" (DARWALL 2013, 16).

[10] Here, I follow CLARKE 2016 and ROSEN 2015 in distinguishing *thoughts* from *beliefs* such that having the latter, but not the former, necessitates assenting to the attitude's propositional content.

[11] For more on this idea, see both CLARKE 2016 (122–123) and ROSEN 2015 (71–72).

[12] Thus, my view is compatible with the possibility of experiencing recalcitrant guilt—that is, with the possibility of experiencing guilt while at the same time believing that one doesn't deserve to feel guilt. My view, then, is a version of what D'Arms and Jacobson (2003) call *quasijudgmentalism*. And although D'Arms and Jacobson raise several worries regarding quasijudgmentalism, I don't find any of them persuasive. And, unfortunately, I don't have the space here to address them all. But take this one. They worry that "the claim that emotions have constitutive thoughts seems incompatible with attributing them to animals and infants, who lack the requisite concepts" (133). Unlike them, I don't find this worrisome. Indeed, although I think that animals and infants are capable of having the same affect, physiological changes, and motivational dispositions that humans typically have when they feel guilt, I doubt that they are capable of feeling guilt precisely because I doubt that they have the requisite concepts.

taking a pill to alleviate one's headache does not. Moreover, this difference isn't just due to the fact that guilt, but not a headache, can be appropriate. For fear can be appropriate and yet there seems to be nothing problematic about taking a pill to rid oneself of one's appropriate fear—except, of course, when it would be instrumentally bad to do so.

It's also important to note that the thought implicated by guilt is not the thought that the wrongdoer deserves to suffer generally.[13] Rather, the thought is only that (1) she deserves to suffer the specific unpleasantness involved in feeling guilty, that (2) she deserves to suffer this unpleasantness only in the right way, at the right time, to the right extent, and with regard to its appropriate object (e.g., her failure to live up to some legitimate demand), and that (3) what's non-instrumentally good is, not her suffering *per se*, but her getting what she deserves (CLARKE 2016). To illustrate, suppose that my wife is accountable for having mistreated me. It would, then, be fitting for me to want her to feel guilty for having mistreated me, and to want this even if this would be of no instrumental value. Nevertheless, it would not be fitting for me to want her to feel lonely. Nor would it be fitting for me to want her to feel guiltier than she deserves to feel.[14] So, again, the idea is only that she deserves to suffer the unpleasantness of guilt and in the right way, at the right time, to the right extent, and with regard to its appropriate object: her failure to live up to some legitimate demand. Moreover, to claim that my wife deserves to feel

[13] Many find this idea unacceptable. For instance, T. M. Scanlon rejects the idea that "it is good that people who have done wrong should suffer" (2013, 102). Likewise, R. Jay Wallace rejects the "problematic thought that wrongdoers positively deserve to suffer" (1994, 108). But rejecting this idea doesn't entail rejecting the idea that someone accountable for some wrongdoing deserves to suffer the unpleasantness of feeling guilty for having committed that wrongdoing. Indeed, Scanlon now accepts that wrongdoers deserve to feel guilty for their wrongdoing—see SCANLON 2008, 188. So, even if we reject the idea that wrongdoers deserve to suffer generally, we shouldn't necessarily reject the idea that wrongdoers deserve to suffer the specific unpleasantness of feeling guilty for their wrongdoing—see, for instance, MCKENNA 2012 (chaps. 6–7)

[14] Why think that she deserves to feel bad at all? Here's my argument: (P1) Given that she's accountable for having mistreated me, it's appropriate to want her to feel guilty for having mistreated me and to want her to have this experience even if her having this experience wouldn't be instrumentally valuable. (P2) If it's appropriate to want *X* even if *X* wouldn't be instrumentally valuable, then *X* must be non-instrumentally valuable. (C1) Thus, her feeling guilty for having mistreated me is non-instrumentally valuable. (P3) What most plausibly accounts for C1 is that she deserves to feel guilty for having mistreated me. (C2) Therefore, she deserves to feel guilty for having mistreated me. And, in defense of P3, I would add, first, that what explains the non-instrumental value of her feeling guilty for having mistreated me is not that her having this experience is itself non-instrumentally valuable. It isn't. After all, her having this experience wouldn't be non-instrumentally valuable if she weren't accountable for her mistreatment of me. Second, the fact that it is *fitting* for her to feel guilty for having mistreated me is not what explains why her feeling guilty is non-instrumentally valuable. For, in general, there's nothing non-instrumentally valuable about having a fitting attitude. There's nothing, for instance, non-instrumentally valuable about fearing what's dangerous even though it is fitting to fear what's dangerous. Thus, it seems that what explains the fact that her feeling guilty is non-instrumentally valuable is both that she deserves to feel guilty and that it is, in general, non-instrumentally valuable that people get what they deserve.

guilty for her mistreatment of me implies only that it would, *in some respect*, be non-instrumentally good that she feels this way, not that her feeling this way would be *overall* good. Thus, the implication is only that the world in which she feels guilty for having accountably mistreated me is, other things being equal, non-instrumentally better than the world in which she likewise feels guilty for having non-accountably mistreated me.[15] This, then, is a very minimal claim about desert; it claims only that feeling appropriate guilt is, other things being equal, less non-instrumentally bad than feeling inappropriate guilt.[16] And, yet, even this rather minimal claim is explanatorily useful. It explains why we think that appropriately feeling guilty is much less lamentable (if lamentable at all) than inappropriately feeling guilty. It explains why it can be appropriate for us to want wrongdoers to feel guilty for what they have done even if the instrumental value of this feeling is, on its own, insufficient to compensate for its unpleasantness. And, in turn, this helps to explain both why we often express our anger and resentment in the hopes of getting those accountable for failing to meet some legitimate demand to feel guilty and why the key to their repairing their relationship with us is for them to feel appropriately guilty for having so failed.

So far, then, I've suggested that a subject is accountable for having φ-ed only if she is the appropriate object of the sorts of reactive attitudes that implicate the thought that she deserves to feel prideful or guilty for having φ-ed. But before I move on to employ this claim in arguing that control is necessary for accountability, I should clarify that, although I have up until now focused mainly on feeling *moral* guilt for immoral *behavior*, I'm not exclusively concerned with either morality or behavior. For we can appropriately feel bad about our imprudent choices, our fallaciously formed beliefs, and our aesthetically distasteful desires (SHOEMAKER 2015, 78). So,

[15] If you think that there's nothing non-instrumentally good about her feeling guilty for having accountably mistreated me, then you would have to think (implausibly) that the world in which she feels guilty for having accountably mistreated me is, other things being equal, no better than the world in which she likewise feels guilty for having non-accountably mistreated me.

[16] Note that my very minimal claim about desert is even more minimal than what others consider to be a relatively minimal claim about desert—see, for instance, Clarke's claim that "if an agent deserves some harm, it will be non-instrumentally good that this harm occurs" (2017, 99). These others are committed to the view that the world in which my wife feels guilty for having accountably mistreated me is, other things being equal, better than the world in which she doesn't feel guilty for having accountably mistreated me. On this view, her feeling guilty is not just, *in some respect*, non-instrumentally good, but is, *overall*, non-instrumentally good. I'm not committed to this stronger claim.

just as we can let others down, we can let ourselves down. Consequently, we may appropriately "beat ourselves up" for having been so stupid, foolish, or distasteful. After all, it's legitimate for us to demand of ourselves that we not be so stupid, foolish, or distasteful. And, so, we can appropriately mentally kick ourselves for such things as swinging at a curveball, laughing at a juvenile joke, and failing to anticipate an obvious objection to an argument. And it strikes us that, in some respect, we deserve to feel bad for having been so stupid, foolish, or distasteful. Consequently, even though we'll certainly lament our having been so stupid, foolish, or distasteful, we won't typically lament our having mentally kicked ourselves for having done so—and this is true even when we don't think that beating ourselves up in this way is instrumentally valuable. Like the feeling of moral guilt, then, feeling bad for having been stupid, foolish, or distasteful strikes us as at least partially deserved. And, so, just as someone can be accountable for her immoral behavior only if she is the appropriate object of the sorts of reactive attitudes that implicate the thought that she deserves to feel guilty for having so behaved, someone can be accountable for, say, her fallaciously formed belief only if she is the appropriate object of the sorts of reactive attitudes that implicate the thought that she deserves to feel bad for having been so stupid. Therefore, we should interpret my use of the word 'guilt' broadly to cover not only moral guilt but also feeling bad for having been stupid, foolish, or distasteful—and, likewise, for how we should interpret my use of the word 'pride'.[17]

I've argued, then, that a subject is accountable for having φ-ed only if she deserves to experience the pleasantness of feeling prideful or the unpleasantness of feeling guilty for having φ-ed. But, intuitively, no subject deserves to experience any feeling on account of her having φ-ed unless φ was under her control.[18] So, we should think that a subject is accountable for having

[17] So, like Gunnar Björnsson (2017), I believe that it can be appropriate to blame ourselves for such things as a poorly considered chess move, an ill-timed pass to a teammate, and a belief formed on the basis of insufficient evidence and that the sort of blame that's appropriate in such instances is not the characteristically moral type that implicates the thought, say, that the agent in question thereby demonstrated ill will. But, unlike Björnsson, I believe that both moral and non-moral blame have retributive elements, at least insofar as it strikes us that we deserve to beat ourselves up for having been so stupid, foolish, or distasteful. We feel that we deserve this, because we have let ourselves down, failing to meet the standards that we legitimately set for ourselves.

[18] I deny, then, what's known as *resultant moral luck* (see ZIMMERMAN 1987)—the idea that one's degree of accountability for φ-ing can be affected by the uncontrolled events that determine the results of one's φ-ing. For some compelling arguments against resultant moral luck, see KHOURY FORTHCOMING.

φ-ed only if φ was under her control. And, to make things absolutely clear, I'll state the argument (the first of two) formally:

(A1.1) For any subject S and any event φ, S is accountable for having φ-ed if and only if it's appropriate for her to feel prideful or guilty for having φ-ed. [Stipulation]

(A1.2) For any subject S, it's appropriate for S to have a given feeling if and only if the thought that's constitutive of that feeling is true. [This is analytic given that 'appropriate' is being used throughout in the alethic sense.]

(A1.3) For any subject S and any event φ, the thought that's constitutive of S's feeling prideful for having φ-ed is the thought that she deserves to experience the pleasantness of feeling prideful, and the thought that's constitutive of S's feeling guilty for having φ-ed is the thought that she deserves to experience the unpleasantness of feeling guilty. [Assumption]

(A1.4) Thus, for any subject S and any event φ, S is accountable for having φ-ed if and only if she deserves to experience either the pleasantness of feeling prideful or the unpleasantness of feeling guilty.[19] [From A1.1–A1.3]

(A1.5) For any subject S and any event φ, S deserves to experience either the pleasantness of feeling prideful or the unpleasantness of feeling guilty only if φ was under her control. [Assumption]

(A1.6) Therefore, for any subject S and any event φ, S is accountable for having φ-ed only if φ was under her control. [From A1.4–A1.5]

The only two assumptions are A1.3 and A1.5. And whereas A1.5 should just seem intuitively obvious, I probably need to say more in defense of A1.3. Now, one reason to think, as A1.3 claims, that the thought that one deserves to feel guilty for having φ-ed is constitutive of feeling guilty for having φ-ed is that it seems impossible to have this feeling without having this thought.[20] Indeed, it seems that this is (at least, in part) what differentiates feeling guilty for

[19] Note that there are a number of people who are willing to accept A1.4 despite not being willing to commit to A1.3. See, for instance, CARLSSON 2017 and CLARKE 2016; Clarke explicitly rejects A1.3, and Carlsson is unwilling to endorse it or any other account of what the thought implicated by guilt is. Of course, all I need is A1.4. So, I offer this argument for A1.4 only to convince those who, unlike Clarke and Carlsson, are not otherwise convinced of A1.4.

[20] When I say that the thought that one deserves to feel guilty for having failed to live up to some legitimate demand is constitutive of feeling guilty, I'm not suggesting that everyone who feels guilty can readily articulate this thought. They can have the thought even if they lack the words to articulate it. Thus, I depart with Rosen in denying that "an account of the thoughts implicit in [a feeling] must be framed in terms that everyone capable of [that feeling] understands" (2015, 75). For although I would

having φ-ed from other ways of feeling bad for having φ-ed (e.g., from feeling ashamed or embarrassed for having φ-ed). For although feeling ashamed or embarrassed for having φ-ed is, like feeling guilty, a way of feeling bad about having φ-ed, feeling guilty seems distinct from these two precisely because one can feel ashamed or embarrassed for having φ-ed without this striking one as being the way that one deserves to feel. And this in turn explains why it seems appropriate to feel ashamed or embarrassed, but not guilty, for things that weren't under one's control. For instance, it seems that it can be appropriate for people to be embarrassed about their physical features or ashamed of their child's behavior even if there was nothing that they could have done to change either their physical features or their child's behavior. For, unlike the thought implicated by feeling guilty, the thoughts implicated by feelings of shame and embarrassment don't implicate the thought that one deserves to feel that way, and, consequently, don't presume that one had control over the intentional object of that feeling.

Another reason to hold that the thought constitutive of feeling guilty for having φ-ed is that one deserves to feel guilty for having φ-ed is that this view (call it *the desert view*) helps us to explain why, after a sufficient amount of time and self-reproach, it ceases to be appropriate to feel guilty anymore. For, on this view, the relevant thought is that one deserves to feel guilty in the right way, at the right time, to the right extent, and with regard to the appropriate object, and this thought (the thought that one still deserves to feel guilty) ceases to be true after a sufficient amount of time and self-reproach has passed/occurred. Thus, after a sufficient amount of time and self-reproach, one ceases to deserve to feel guilty anymore. And, consequently, it ceases to be appropriate to feel guilty.

To better understand these advantages for the desert view, it will be helpful to contrast it with the following three alternatives: (1) the quality-of-will view, according to which the thought implicated by guilt is that the agent's conduct manifests ill will, (2) the blameworthy view, according to which the thought implicated by guilt is that the agent is blameworthy, and

agree that they must have the relevant concepts, I believe that someone can have the relevant concepts without understanding the terms that we would use to articulate them. For instance, a young child who has seen pictures of griffins, lions, and eagles as well as pictures of their various parts could have the concept of a griffin without being able to understand all or any of the terms used in the following articulation: a creature that has the torso, tail, and back legs of a lion and the head and wings of an eagle.

(3) the wrong-doing view, according to which the thought implicated by guilt is that one has done wrong. I'll take each in turn.

Consider, first, the quality-of-will view. For one, it seems possible to feel guilty without having the thought that one has manifested ill will—that is, without it even seeming that one has acted out of ill will. For instance, a woman could feel guilty (although inappropriately so) for running over a boy who unexpectedly darted in front of her car without it seeming to her that her conduct manifested ill will. Perhaps, she realizes that it was physically impossible for her to have avoided running over the boy given the way that he unexpectedly darted in front of her car. Still, it could be guilt that she's feeling so long as her feeling bad about having run over the child strikes her as (at least, partially) deserved.[21]

For another, if the quality-of-will view were right about the thought that was constitutive of feeling guilty, there would be no reason to think that it would, after a sufficient amount of time and self-reproach, cease to be appropriate for one to feel guilty for having previously manifested ill will. For no matter how much time has passed and no matter how much one has already reproached oneself for that previous lapse, it will never cease to be true that one did manifest ill will. And, so, it will never cease to be appropriate to feel guilty for having done so. Yet, intuitively, it seems that, after a sufficient amount of time and self-reproach, one should cease to rebuke oneself for one's past failures.

Next, consider the blameworthy view. Here, too, it seems possible for someone to feel guilty without having the thought that one is blameworthy. For it seems that one needn't even have the concept of blameworthiness to feel guilty. And this makes sense of our thought that children can feel guilty even if they lack the concept of blameworthiness. This is because, as anyone with much experience with children knows, children have the concepts of merit, desert, and fairness, which is all that, on the desert view, is needed to feel guilt. Thus, it seems that what's needed to feel guilty is, not the concept of blameworthiness, but the concept of merit, desert, or fairness.

[21] In this case, it will likely strike her as if she deserves to feel guilty for killing the boy, but this is as much an illusion as a straight stick half immersed in water striking her as bent. Such illusions arise when the apparatus that we use for making such cognitions have been trained, or have evolved, under a range of contexts that differs significantly from the one in which the illusion presents itself. (I thank Mark Schroeder for reminding me of this.)

And, again, we see that, like the quality-of-will view, the blameworthy view has trouble explaining why, after a sufficient amount of time and self-reproach, it ceases to be appropriate to feel guilty anymore. For, again, the relevant thought—which, in this case, is the thought that one was blameworthy for some past misdeed—never ceases to be true. And, so, we must on this view counterintuitively hold that no matter how much time has passed and no matter how much one has already reproached oneself for one's previous failure, it never ceases to be appropriate to feel guilty for that failure.

Lastly, consider the wrong-doing view. Here, too, it seems that one can feel guilty for having φ-ed without thinking that one has done wrong in having φ-ed. To illustrate, imagine that Huck Finn (Mark Twain's famous fictional character) had, contrary to the actual story, turned his friend Jim over to the authorities for being a runaway slave. In such a case, it seems possible for Huck to feel guilty for having turned his friend in even if it doesn't strike him as wrong for him to have done so. For so long as it strikes him that he deserves to feel guilty for doing so, it seems to be guilt that he's feeling. And, so, we should reject the wrong-doing view, because it wrongly assumes that we must have thoughts of having done wrong to feel guilty. What's more, we should reject the wrong-doing view, because, like the other two alternatives, it fails to account for the fact that, after a sufficient amount of time and self-reproach, it ceases to be appropriate to feel guilty anymore. And it fails for the exact same reason that the other two failed: the thought in question is not one that ever ceases to be true.[22]

To sum up, then, we should accept that a subject can be accountable for only that which was under her control. For, according to the desert view, a subject can be accountable for having φ-ed only if she deserves to experience the relevant feeling (pride or guilt) for having φ-ed. And no one, it seems, deserves to experience any feeling on account of one's φ-ing when φ wasn't under one's control. But, of course, some would argue that things are not as they seem. That is,

[22] I'm assuming that, on the wrong-doing view, what's implicated by one's feeling guilty for having φ-ed is the thought that it was all-in wrong for one to have φ-ed. If it were instead the thought that it was merely *pro tanto* wrong for one to have φ-ed, then it would, on this account, be appropriate for one to feel guilty for having broken a relatively unimportant promise so as to save a life. But this is inappropriate despite the fact that it's *pro tanto* wrong to break a promise.

some challenge the idea that subjects can be accountable only for that over which they exerted control, and they do so by appealing to Frankfurt-type cases, such as the following.[23]

> *The Sniper:* Jones is a sniper who plans to shoot the mayor when she takes the podium at noon today. Black also wants the mayor dead and believes that Jones will decide on his own to pull the trigger, but he doesn't want to take any chances. So, he implants a deterministic device in Jones's brain that will issue, at noon, in a decision to kill the mayor if Jones's own cognitive faculties aren't going to issue, at noon, in a decision to pull the trigger. Moreover, Jones will succeed in shooting the mayor if either his own cognitive faculties or Black's device issues in a decision to pull the trigger. As it happens, Jones's own cognitive faculties issue in a decision to pull the trigger, and, as a result, the mayor is shot dead. But neither Black's device nor its implantation played any causal role in the killing of the mayor. Things proceeded just as they would have had neither Black nor his device existed.

In this case, it certainly seems appropriate for Jones to feel guilty—or, at least, it seems that way so long as we are to imagine that Jones is a normal person with a normal history and, thus, not a psychopath or someone who has always lived under Black's influence. But even if it's clear that it's appropriate for Jones to feel guilty, it's not at all clear that we can legitimately hold him accountable for having killed the mayor. Indeed, there's good reason to think that we can't. After all, we're to assume that he didn't have the option of refraining from killing her. Indeed, what makes this a Frankfurt-type case is that he didn't have this option. But, now, the nature of options is such that Jones was under an obligation to refrain from killing her only if this was an option for him, which it wasn't. For I'll just stipulate that, as I'll use the term, an option for a subject is any member of the set such that, for any possible event φ, whether her φ-ing has a deontic status (e.g., that of being obligatory) depends on whether φ is a member of this set; and, if it is a member, what particular deontic status it has depends on how it compares to the other members in this set.[24] Thus, given that he lacked the option to refrain from killing

[23] I'm presuming that the relevant sort of control is what's known as *regulative control* as opposed to *guidance control*, the key difference being that you can have the latter, but not the former, with respect to φ even if not–φ-ing isn't an option. (For more on this distinction, see FISCHER & RAVIZZA 1998.)

[24] Given this stipulation, there can be no question that 'one is obligated to φ' implies 'one has the option of φ-ing' even if there is some question whether 'one is obligated to φ' implies 'one can φ'. For more on this, see PORTMORE FORTHCOMING.

the mayor, he couldn't have been under any obligation to refrain from killing her. And if that's right, I don't see how we can think that Jones deserves to feel guilty for having killed her. Why would he deserve to suffer the unpleasantness of guilt for having done something permissible? Moreover, if he didn't have the option of doing anything other than kill the mayor, what could we have legitimately expected him to have done instead?[25]

So, given both the necessary connection between being blameworthy (or otherwise accountable) for having φ-ed and having had an obligation to refrain from φ-ing and the necessary connection between having had an obligation to refrain from φ-ing and having had the option to refrain from φ-ing, we should think that a subject can be blameworthy (or otherwise accountable) for having φ-ed only if she had the option to refrain from φ-ing. And since Jones didn't have the option to refrain from killing the mayor, he can't be blameworthy for having done so.

Here, then, is my second argument for control being necessary for accountability.[26]

(A2.1) For any subject S and any event φ, S is accountable for having φ-ed if and only if S is praiseworthy/blameworthy for having φ-ed. [Stipulation]

(A2.2) For any subject S and any event φ, S is praiseworthy/blameworthy for having φ-ed only if she ought to have φ-ed/was obligated to refrain from φ-ing.[27] [Assumption]

---

[25] This is Widerker's "What-should-he-have-done defense" (called the "W-defense" for short) of the principle of alternative possibilities. See WIDERKER 2000.

[26] This argument is inspired by those of COPP 1997 and WIDERKER 1991. One key difference is that whereas Copp and Widerker rely on the controversial principle that 'ought' implies 'can', I rely on the incontrovertible principle that both 'ought' and 'obligation' imply 'option'. This principle is incontrovertible, for I've defined an 'option' such that only options are eligible for deontic status—see note 24. The other key difference is that whereas they focus exclusively on blame, my argument is about accountability in general and, thus, concerns praise as well as blame.

[27] Ishtiyaque Haji (1997) endorses a version of A2.2, which he calls "the Objective View of Blameworthiness." On this view, if a subject is blameworthy for φ-ing, then it must be that she had an obligation to refrain from φ-ing. And this view is presupposed by many others—see, for instance, COPP 1997 and PORTMORE 2011 (chap. 2). Also, note that, for an act to be praiseworthy, it must have been something that one ought to have performed, not merely something that one was permitted to perform. By contrast, it can be that one ought to have done better without one's being blameworthy for not having done better. This is because an act is blameworthy only if it is contrary to what one was obligated to do. Thus, one can be blameworthy for not having given more to charity only if one was *obligated* to give more. That one *ought* to have given more is insufficient to make one blameworthy.

Interestingly, given that it's a conceptual truth that 'one is obligated to φ' implies 'one has the option of φ-ing', Fischer must bite the bullet and accept both that nothing anyone ever does is morally wrong and that, nevertheless, some people are morally blameworthy for doing what they were permitted to do—see FISCHER 2003 (esp. 249). Clearly, then, Fischer has a much higher tolerance for bullet-biting than I do.

(A2.3) For any subject S and any event φ, S ought to have φ-ed/was obligated to refrain from φ-ing only if she had the option of φ-ing/refraining from φ-ing. [Analytic given my stipulative definition of 'option'.]

(A2.4) For any subject S and any event φ, S had the option of φ-ing/refraining from φ-ing only if whether she was to φ was under her control. [Assumption]

(A2.5) Therefore, for any subject S and any event φ, S is accountable for having φ-ed only if whether she was to φ was under her control. [From A2.1–A2.4]

Of course, many would view *The Sniper* as a counterexample to A2.2.[28] They would claim that Jones is blameworthy for killing the mayor even though he had no obligation to refrain from doing so given his lack of the option of so refraining. But although it's clear that Jones is blameworthy, it's not clear that he is directly blameworthy for *killing the mayor*. And those who would deny A2.2 must assume that this is what he is directly blameworthy for. Yet, it's plausible to suppose that he's instead directly blameworthy for one or more of the following: (1) having formed the malicious desire to kill the mayor;[29] (2) having formed a willingness to do what he believes to be wrong; (3) having killed the mayor *on his own accord* or *for his own reasons* (NAYLOR 1984, ROBINSON 2012, and WEDGWOOD 2017); or (4) having exercised his rational capacities in a way that didn't trigger Black's device (WEDGWOOD 2017). I think that the only reason some may resist such alternatives is that they assume that Jones never had the relevant sort of control over any of 1–4 and, so, can't be accountable for any of them.[30] But, in the next section, I'll argue that although Jones didn't have volitional control over any of 1–4, he

[28] Others take the following to be a counterexample to A2.2. Suppose that Arthur, a white supremacist, sneaks up behind an unsuspecting black man, named Bert, and clubs him over the head, knocking him unconscious. He does so out of a hatred of blacks. However, unbeknownst to Arthur, Bert was just about to shoot his ex-girlfriend Carla, who, we'll suppose, was completely innocent. As it turns out, then, Arthur's act saves an innocent life. Yet, Arthur is clearly blameworthy. So, this may seem to be a counterexample to A2.2. But it's a counterexample only if we think that Arthur was obligated to refrain from clubbing Bert. And we shouldn't think this. For if we thought this, we have to think (implausibly) that it would be appropriate for us to demand that he refrain from clubbing him, which it wouldn't be. Instead, we can rightly demand only that he not do so out of racial hatred. Thus, when we realize what Bert was actually blameworthy for (that is, his bad motive), we find that this is no counterexample to A2.2.

[29] You may wonder: what if we were to imagine a variant on this case in which there was an additional device that would have caused Jones to form the desire kill the mayor had he not done so via his own cognitive faculties. In that case, I don't see how we could legitimately hold him accountable for forming this malicious desire. For, in that case, he lacked the option of not forming this malicious desire. So, I think that all that we can hold him accountable for is his forming this malicious desire via his own cognitive faculties.

[30] See, for instance, FISCHER 1994 (143–145).

did have the relevant sort of control over all of 1–4. And this, I think, will defuse whatever force that such a Frankfurt-type case is meant to have. For, in that case, we can account for our intuition that Jones is deserving of blame without our having to reject our intuition that a subject can be accountable for having φ-ed only if she had an obligation to refrain from φ-ing.[31] So, let me now turn to explaining what I take the relevant sort of control to be.

### 3. The Third Horn: Deny that the Relevant Sort of Control is Volitional Control

As I've noted, we don't typically, if ever, exert volitional control over our reasons-responsive attitudes. This along with the common intuition that people can both be obligated to form certain reasons-responsive attitudes and be accountable should they fail to do so has led several philosophers to grab the third horn and argue that the sort of control that's relevant to determining our obligations and responsibilities isn't volitional control, but *rational control*, where this is the sort of control that we exercise by being both receptive and reactive to reasons—forming, revising, sustaining, and/or abandoning our reasons-responsive attitudes in light of our awareness of facts (or what we take to be facts) that count for or against them.[32]

This, I believe, is the right way to go, but, to assess the plausibility of this approach, we need to better understand what rational control is. The following is a tentative proposal, and,

---

[31] So, I would deny that the lesson of Frankfurt-type cases is that we should deny that control is necessary for accountability. Instead, I take the lesson to be that what we're directly accountable for is, not our voluntary actions, but how we exercise our rational capacities and whether we do so in a way that fails to trigger Black's device.

[32] The term 'rational control' comes from McKenna 2012, A. M. Smith 2005, and A. M. Smith 2015. Others use a different term for the sort of control that we exert directly over our reasons-responsive attitudes. For instance, Pamela Hieronymi (2006) uses the term 'evaluative control', Conor McHugh (Forthcoming) uses the term 'attitudinal control', and Ralph Wedgwood (2017) uses the term 'deliberative control'. And each of these philosophers has a slightly different idea about what exactly this sort of control amounts to. But the important thing is that they all agree that there's a kind of control that we exert directly over our reasons-responsive attitudes.

Now, some may question whether control that doesn't involve the exercise of one's volitions counts as genuine control. But, to my mind, this begs the question. For, in ordinary English, to exercise control over something is just to manage, regulate, or influence it. Consequently, it's perfectly felicitous to talk of our heartrate being controlled by our autonomic nervous system, of our engine's air-fuel ratio being controlled by its carburetor, and of our office's temperature being controlled by its thermostat. Of course, we should arguably think that the relevant sort of control must be the sort of personal control that, for instance, I exert over the contractions of my biceps when I flex them while looking at myself in the mirror as opposed to the sort of sub-personal control that I exert over the contractions of my cardiac muscles when their frequency increases in response to my increased anxiety. But, as I argue in chapter 3 of Portmore Forthcoming, rational control is indeed a kind of personal control.

although it's somewhat complicated, it basically just says that for φ to be under one's rational control is for it to depend on whether, and how, one exercises one's rational capacities.

> **Rational Control:** For any event φ, a subject has, as of time *t*, rational control over whether she φs at some later time *t'* if and only if she has, as of *t*, the relevant rational capacities and whether she φs at *t'* depends (and in the right way) on whether, and how, she exercises these capacities at *t*. And a subject has, as of time *t*, the relevant rational capacities with respect to whether she φs at some later time *t'* if and only if she is inherently so structured that, in a suitably wide range of possible worlds, she recognizes the considerations that count for and against her φ-ing at *t'* and, consequently, φs or refrains from φ-ing at *t'*, depending on which these considerations make appropriate.[33]

To illustrate, consider *Stupid Mistake:* a genius named Albert takes a math test and misses one of the easiest problems because he overthinks things and, consequently, overlooks its simple solution.[34] In this case, Albert had rational control over whether he would provide the correct solution to the problem. After all, he was, we'll assume, inherently so structured that he would have come up with the correct solution in a suitably wide range of possible worlds. For let's assume that, had he used his rational capacities to stop and think about what the solution might be like, he would have recognized the possibility that it could be quite simple. And let's assume that, had he recognized this possibility, he would have come up with the correct solution.

Of course, the reader may wonder what sorts of worlds must a suitably wide range of possible worlds include? On this issue, I'll remain neutral. But if incompatibilism is true, then such a range must include worlds in which Albert exercises his rational capacities differently so as to comes up with the correct solution to the problem even though the causal laws and histories of these worlds are identical to those of the actual world.[35] But if, instead,

---

[33] This account of rational capacities is inspired by Michael Smith's account—see his 2003.

[34] This is partly inspired by John Maier's example of a professional golfer who misses what should have been a gimme putt. Maier (2014) uses it to question whether its being the case that S would φ if she were to try to φ is necessary for S's having the option to φ. The example originates with AUSTIN 1956. My example, though, is also partly inspired by Michael Smith's example of someone blanking on the answer to a question—see his 2003.

[35] Indeed, some incompatibilists may want to claim that the sort of control that's required for accountability is incompatible with determinism and that, therefore, what I'm calling 'rational control' will be of the right sort only if we understand the phrase

compatibilism is true, then such a range needn't include such worlds. For, if compatibilism is true, such a range need only include worlds in which Albert's brain is structured exactly as it is in the actual world and yet he comes up with the correct solution because the casual history of this world differs slightly and in such a way that he exercises his rational capacities differently, stopping to think about what the solution might be like and, consequently, recognizing the possibility that it might be quite simple.[36] Thus, my account of rational control allows us to make sense of the compatibilism-incompatibilism debate. The debate concerns whether rational control, appropriately understood, is compatible with causal determinism, which, in turn, depends on whether its notion of a suitably wide range of possible worlds must include ones in which Albert comes up with the correct solution even though the causal laws and histories of these worlds are identical to those of the actual world in which he doesn't come up with the correct solution. And I take it to be a merit of my view that it makes room for such debate.

To better understand my account of rational control, it will be helpful to contrast the above case with one in which a toddler named Todd fails to provide the correct solution to the same problem. In this variant, we should deny that Todd had rational control over whether he was to provide the correct solution. For even though the math problem is a relatively easy one for a math genius, we can plausibly assume that in no nearby possible world in which we hold fixed the way that his underdeveloped brain is structured does Todd come up with the correct solution to the problem. For, in the case of the toddler, there is no nearby possible world in which we simply vary the way he exercises his rational capacities (e.g., exercising them so as to consider all kinds of possible solutions) and Todd comes up with the correct solution. In this case, we can account for Todd's failure to come up with the correct solution only by appealing

'suitably wide range of possible worlds' to include worlds in which the agent acts differently even though the causal laws and histories of these worlds are identical to those in the actual world. But, again, I'll remain neutral on such issues.

[36] For some doubts about whether we could ever be accountable on the basis of some "fluke" such as our casual history just happening to be one way rather than another, see MCGEER & PETTIT 2015. But, perhaps, we can be accountable for how our rational capacities operate even in circumstances in which their operation is determined by some such "fluke" in our causal history. For, perhaps, it's sufficient that we've taken responsibility for the operation of our rational capacities in such circumstances in that we've come to accept that it's "fair," in the sense of being part of our given social practices, that we be subject to reactive attitudes in virtue of how our rational capacities operate in such circumstances. For more on the relevant notion of 'taking responsibility', see FISCHER & RAVIZZA 1998. And, for more on how we can be responsible for something (such as a "glitch" or an implicit bias) in virtue of our having taken responsibility for the mechanism that issued in that something, see MASON 2018.

to his lack of certain capacities, and not by appealing to his failure to exercise those capacities appropriately.

Of course, the reader may further wonder: If we're to take a compatibilist approach, to what degree and in what ways are we allowed to vary the causal history (or laws) while holding the brain's structure fixed? Here, too, I'll remain neutral. For the only points that I wish to make are that (1) there seems to be a perfectly ordinary sense in which Albert could have come up with the correct answer but Todd couldn't and that (2) by appealing to the presence or absence of rational control we are able to explain this difference in terms of the differences in the rational capacities of the two subjects. So, whereas Todd failed to come up with the correct answer because he lacked the relevant capacities, Albert failed to do so because he failed to exercise those capacities appropriately. It seems, then, that the way to determine whether a subject had the option of doing other than what she did is to look, first, at whether she had the relevant rational capacities. And, if she did, to look, second, at whether the best explanation for her not doing other than what she did is either *her* failure to exercise her capacities appropriately or someone or something else preventing her from doing so.[37] Thus, if, in *Stupid Mistake*, Black had been standing ready to intervene had Albert shown any sign of being about to exercise his rational capacities appropriately, the best explanation for Albert's failure to come up with the correct solution would not have been that *he* failed to exercise his capacities appropriately, but that *Black* foreclosed the possibility of his doing so. And, in such a case, Albert could be accountable only for having failed to trigger Black's intervention, not for having failed to come up with the correct solution.

So, if we hold that the relevant sort of control is rational control as opposed to volitional control, we can accommodate our intuition that people are directly accountable for their reasons-responsive attitudes, and we can do so without implausibly denying that control is necessary for accountability. Moreover, we can account for our intuitions in Frankfurt-type cases by taking the subject in question to be accountable for having failed to exercised her

[37] Here, I follow M. SMITH 2003.

rational capacities in such a way as to trigger Black's device and not for having failed to do what the device made it impossible for her to do, which is implausible.

**4. Conclusion**

Intuitively, it seems that we can be directly accountable for our reasons-responsive attitudes—e.g., our beliefs, desires, and intentions. And, yet, we rarely, if ever, have volitional control over such attitudes. This presents a trilemma: (Horn 1) deny that we can be directly accountable for our reasons-responsive attitudes, (Horn 2) deny that φ's being under our control is necessary for our being directly accountable for φ-ing, or (Horn 3) deny that the relevant sort of control is volitional control. I've argued that we should take hold of Horn 3. Of course, if it turns out both that determinism is true and that determinism is incompatible with our having rational control over our reasons-responsive attitudes (a possibility that I've allowed for), we'll be forced to take hold of Horn 1 as well. But, in that case, we're probably not accountable for anything. For it seems that we can be accountable for our actions only if we have voluntary control over them, which involves both volitional control over the actions and rational control over the volitions that give rise to them. So, if we don't have rational control over our reasons-responsive attitudes, we won't have rational control over our volitions. And if we don't have rational control over our volitions, we won't be accountable for our actions. And if we're accountable neither for our reasons-responsive attitudes nor for our voluntary actions, it seems that we're accountable for nothing. The upshot of this paper, then, is that the case for our being accountable for our reasons-responsive attitudes is no worse than that for our being accountable for our voluntary actions and that, if it turns out that we're not accountable for our reasons-responsive attitudes, we're probably accountable for nothing.[38]

[38] I thank Gunnar Björnsson, Andreas Brekke Carlsson, Yishai Cohen, David Copp, Liz Harman, Brian Hedden, Hrishikesh Joshi, Andrew Khoury, Michael McKenna, Berislav Marušić, Ellie Mason, Conor McHugh, Philip Pettit, Susanna Rinard, David Shoemaker, David Sobel, Daniel Star, Steve Sverdlik, Matthew Talbert, Travis Timmerman, and audiences at UC Davis, Princeton University, Syracuse University, University of Arizona, Arizona State University, Australian National University, University of Colorado at Boulder, University of Maryland at College Park, the 2014 Pacific Division Meeting of the American Philosophical Association, and the 4th biennial New Orleans Workshop on Agency and Responsibility.

# References


Austin, J. L. (1956). "Ifs and Cans," *Proceedings of the British Academy* 42: 107–132.

Björnsson, G. (2017). "Review of Rik Peels, Responsible Belief: A Theory in Ethics and Epistemology." *Notre Dame Philosophical Reviews*.

Carlsson, A. B. (2017). "Blameworthiness as Deserved Guilt." *Journal of Ethics* 21: 89–115.

Chappell, R. Y. (2012). "Fittingness: The Sole Normative Primitive." *The Philosophical Quarterly* 62: 684–704.

Clarke, R. (2016). "Moral Responsibility, Guilt, and Retributivism." *Journal of Ethics* 20: 121–137.

———. (2013). "Some Theses on Desert." *Philosophical Explorations* 16: 153–164.

Copp, D. (1997). "Defending the Principle of Alternate Possibilities: Blameworthiness and Moral Responsibility." *Noûs* 31: 441–56.

Darwall, S. (2013). "Morality's Distinctiveness." In S. Darwall (ed.) *Morality, Authority, and Law: Essays in Second-Personal Ethics I*, pp. 3–19. Oxford: Oxford University Press.

D'Arms, J. and D. Jacobson (2003). "The Significance of Recalcitrant Emotion (or, Anti-Quasijudgmentalism)." *Philosophy* 52 (Supp): 127–145.

———. (2000). "The Moralistic Fallacy: On the 'Appropriateness' of the Emotions." *Philosophy and Phenomenological Research* 61: 65–90.

Fischer, J. M. (2003). "'Ought-Implies-Can', Causal Determinism and Moral Responsibility." *Analysis* 63: 244–250.

———. (1994). *The Metaphysics of Free Will: An Essay on Control*. Oxford: Wiley-Blackwell.

Fischer, J. M., and M. Ravizza. (1998). *Responsibility and Control: A Theory of Moral Responsibility*. Cambridge: Cambridge University Press.

Frankfurt, H. G. (1969). "Alternate Possibilities and Moral Responsibility." *The Journal of Philosophy* 66: 829–839.

Haji, I. (1997). "Frankfurt-Pairs and Varieties of Blameworthiness: Epistemic Morals." *Erkenntnis* 47: 351–377.

Hieronymi, P. (2006). "Controlling Attitudes." *Pacific Philosophical Quarterly* 87: 45–74.

Khoury, A. C. (Forthcoming). "The Objects of Moral Responsibility." *Philosophical Studies*.

Maier, J. (2014). "Abilities." In E. N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Fall Edition), URL = <http://plato.stanford.edu/archives/fall2014/entries/abilities/>.

Mason, E. (2018). "Respecting Each Other and Taking Responsibility for Our Biases." In K. Hutchison, C. Mackenzie, and M. Oshana (eds.), *Social Dimensions of Moral Responsibility*, pp. 163–184. Oxford: Oxford University Press.

McGeer, V. and P. Pettit (2015). "The Hard Problem of Responsibility." In D. Shoemaker (ed.) *Oxford Studies in Agency and Responsibility, Volume 3*, pp. 160–188. Oxford: Oxford University Press.

McHugh, C. (2017). "Attitudinal Control." *Synthese* 194: 2745–2762.

McKenna, M. (2012). *Conversation and Responsibility*. Oxford: Oxford University Press.

Mill, J. S. (1991). [1861]. *Utilitarianism*. In J. M. Robson (ed.), *Collected Works of John Stuart Mill*. Vol. 10, pp. 203–59. London: Routledge.

Morris, H. (1976). "Guilt and Suffering." In his *On Guilt and Innocence: Essays in Legal Philosophy and Moral Psychology*, pp. 89–110. Berkeley: University of California Press.

Naylor, M. (1984). "Frankfurt on the Principle of Alternate Possibilities." *Philosophical Studies* 46: 249–258.

Portmore, D. W. (FORTHCOMING). *Opting for the Best: Oughts and Options*. New York: Oxford University Press.

———. (2011). *Commonsense Consequentialism: Wherein morality meets rationality*. New York: Oxford University Press.

Robinson, M. (2012). "Modified Frankfurt-type Counterexamples and Flickers of Freedom." *Philosophical Studies* 157: 177–194.

Rosen, G. (2015). "The Alethic Conception of Moral Responsibility." In R. Clarke, M. McKenna, and A. M. Smith (eds.) *The Nature of Moral Responsibility: New Essays*, pp. 65–87, Oxford: Oxford University Press.

Scanlon, T. M. (2013). "Giving Desert Its Due." *Philosophical Explorations* 16: 101–116.

———. (2008). *Moral Dimensions. Meaning, Permissibility and Blame*. Cambridge: Harvard University Press.

Shoemaker, D. (2015). *Responsibility from the Margins.* Oxford: Oxford University Press.

Smith, A. M. (2015). "Attitudes, Tracing, and Control," *Journal of Applied Philosophy*. Advance online publication available at http://dx.doi.org/10.1111/japp.12107.

———. (2005). "Responsibility for Attitudes: Activity and Passivity in Mental Life," *Ethics* 115: 236–271.

Smith, M. (2003). "Rational Capacities, or: How to Distinguish Recklessness, Weakness, and Compulsion." In S. Stroud and C. Tappolet (eds.), *Weakness of Will and Practical Irrationality*, pp. 17–38, Oxford: Clarendon Press.

Wallace, R. J. (1994). *Responsibility and the Moral Sentiments*. Cambridge: Harvard University Press.

Watson, G. (1996). "Two Faces of Responsibility." *Philosophical Topics* 24: 227–248.

Wedgwood, R. (2017). *The Value of Rationality*. Oxford: Oxford University Press.

Widerker, D. (2003). "Blameworthiness and Frankfurt's Argument Against the Principle of Alternative Possibilities." In D. Widerker and M. McKenna (eds.), *Moral Responsibility and Alternate Possibilities*, pp. 265–99. Burlington, VT: Ashgate, 2003.

———. (2000). "Frankfurt's Attack on Alternative Possibilities: A Further Look." *Philosophical Perspectives* 14: 181–201.

———. (1991). "Frankfurt on 'Ought Implies Can' and Alternative Possibilities." *Analysis* 51: 222–24.

Zimmerman, M. J. (1987). "Luck and Moral Responsibility." *Ethics* 97: 374–386.