

## Character, Will, and Agency

Roman Altshuler

[Forthcoming in *From Personality to Virtue: Essays on the Philosophy of Character*, eds. Jonathan Webber and Alberto Masala (Oxford University Press, 2016). This is the author's copy and may differ from the published version.]

Abstract: Character and the will are rarely discussed together. At most, philosophers working on the one mention the other in an eliminativist vein—if character is represented as something chosen, for example, it can be chalked up to the work of the will; if the will consists merely of a certain arrangement of mental states, it can be seen as little more than a manifestation of character. This mutual neglect appears perfectly justified. If both character and will are determinants of action, to treat them separately would be to overdetermine agency at best, and at worst to fragment it. While defending this reasoning, I argue that things are not so simple, because character and will serve distinct explanatory and normative functions, respectively. The difference in function, however, does not prove that character and will must be ontologically distinct sources of agency; only that our discourse about them must keep them apart.

Character is commonly understood as a set of dispositions to respond in certain ways to one's surroundings, and as such is a broad category. It can include, for example, dispositions to feel or to react to events, or even more broadly, dispositions to see particular features of one's surroundings as salient and calling for action, thus playing a role in motivating action as well as moral perception. But aside from character, it seems, we also have a will, a capacity for making decisions, and the relation between these two features of agency is frequently overlooked. I will be concerned first to suggest that there are good reasons to think there is a tension between holding that both character and the will have a role to play in human agency, and that this tension inclines supporters of one toward eliminativism about the other. But

I will also argue that eliminativism is inadvisable: neither character nor the will can be eliminated from our understanding of agency without a cost, because each serves functions for which the other is unsuited. However, I will aim to cast doubt on the idea that this consideration genuinely implies that will and character must be distinct phenomena; instead, acting volitionally and acting from character (for lack of better terminology) may provide different descriptions of the same events. This will require modifications to the standard views of both will and, to some extent, character. Finally, however, I will argue that this should not be taken to mean that we can speak of will and character interchangeably or, conversely, that we should be free to do away with one of them. Character talk and will talk—and the self-conceptions that go along with such talk—serve important and irreducible purposes.

### **I. Toward a Single Theory of Agency**

Authors focusing on character tend to avoid talk of the will<sup>1</sup>; similarly, those concerned with the will frequently avoid character, unless they are attempting to eliminate the former by reducing it to the latter. This eliminativism makes perfect sense, on both sides. Character, after all, is supposed to determine our agency. While the dispositions involved in character may be dispositions to feel, reason, or desire in particular ways, they are first and foremost dispositions to act. In his discussion of the virtues of character, Aristotle, for example, notes that virtue involves having the right knowledge, the right motive, and the right disposition toward one's passions, but no one can accomplish the human function without virtuous

---

<sup>1</sup> There are, of course, apparent exceptions. But many are not genuine exceptions to the rule. For example, Hudson (1980) argues that understanding character traits requires attention to the role of strength and weakness of will: being virtuous, on his view, involves not simply acting rightly or having the right motives, but acting rightly because one wills to act on those motives. But strength of will is itself a character trait, so that the notion of will referenced here is fully subservient to that of character.

*action*. Production of such action is, in other words, the *point* of virtue. And, as many writing on character have pointed out, character is intimately tied to prediction of behavior: to know someone's character is to know, or at least be in a position to make strong predictions about the person's future behavior.

With the will, however, we find ourselves in a similar position. If an action is to be attributable to me as an agent, it must be because I somehow chose or decided to carry it out. Were my arm to simply jerk as a result of an electrical current applied to my nerves, the resulting movement would not be an action at all but simply a spasm. So our acts of will, our volitions, however we choose to interpret them, must explain why some movement of my body took place if that movement is to be an action. And now it seems that we have a problem, since both character and will seem to play the same theoretical role: they explain why an action took place. But if actions are explicable in terms of both character and will, it begins to look as if actions are overdetermined. Why, after all, would we *need* two distinct ways of explaining why an action took place?

It does not help here to point out that on most currently dominant accounts actions are explicable in terms of some combination of beliefs and desires, since those alone cannot explain why an action took place. Even if we grant, and this is far from a given, that any time we act there must be (at a minimum) some belief-desire pair behind the action, the belief-desire pair by itself is not sufficient to explain why we performed that action rather than some other, given that there are often a number of distinct ways to satisfy a given desire and belief (Goldie 2004a), nor, for that matter, why we performed that action at all, since we have many desires we never act on. Even if we grant, as Michael Smith (2012) holds, that citing beliefs and desires places actions within a familiar context of explanation, questions about why the agent had those beliefs and desires, or about why she acted on

them, remain. The belief-desire explanation is incomplete unless it is filled out with either character or will.

On one hand, we might explain how the agent saw the situation and why seeing it in this way gave rise to certain desires—this is a type of explanation in terms of character (Butler 1988). As Annette Baier notes in her incisive paper on Hume and character, we can often cite beliefs or desires of the agent as explanation because so long as these are themselves ordinary enough, reference to them will suffice to make the action intelligible. But “when either the motivation or the beliefs of the agent themselves cry out for explanation...citing them does not satisfy our wish to understand, nor end our puzzlement. Then a description of the agent can help,” because “[t]he kind of person one is helps determine both the sort of desires one has and acts on, and the way one selects advisors and forms one’s beliefs” (Baier 2009, 246). So belief-desire pairs are, on this account, merely steps in the transition from character to action. While Baier herself rejects the implication that all action explanations must cite character, this is ultimately due to her view of explanation: she takes an explanation to be sufficient so long as it makes the explanandum intelligible, something explanations in terms of belief-desire pairs can usually accomplish when the beliefs and desires themselves are intelligible enough. That we often *can* make an action intelligible without reference to character, however, does nothing to show that a certain understanding of character is not already presupposed in that intelligibility; many commonplace actions are, after all, perfectly intelligible even without reference to any beliefs or desires of the agent, but presumably no defender of the belief-desire theory would be willing to dispose of mental states in such cases.

Alternatively, we might note that the agent *chose* to act on this particular desire rather than some other in these circumstances, and this will provide us with an explanation in terms of the will. For, a defender of the will might

argue, desires do not simply push us into particular actions, but must first be in some way endorsed by the agent. An agent might take acting on the desire to be more intelligible than the alternative, for example (Velleman 1992), or as providing an answer to the question of what to do (Hieronymi 2009); here the will as a capacity for practical reason steps in to fill out the belief-desire conception. On other views, the idea of the will as a separate executive capacity may be brought in to make sense of the choice among desires. Perhaps, the will manufactures desires of its own; on one view, motivational states can lead to intentional action only if they are themselves produced by and under the direct control of the will, so that those states are better characterized as choices or decisions than simple desires, to which we are passively subject (Wallace 2006). In any case, however, if *either* a volitionist or characterological account can be brought in to explain action, agency will appear overdetermined. Attempts to avoid this problem of overdetermination lead to a tendency in the direction of taking either character or the will as providing the *ultimate* explanation, to which the other is subservient. But of course if we can explain an event in terms of one factor, which is itself explicable in terms of another, the instability quite naturally leads toward an attempt to eliminate the subservient explanation.

There is a further way to skirt the issue: one might simply insist that character does not determine action at all. One version of this view, for example, might hold that our formative experiences, those that have shaped our character, are not sufficient to determine our actions in every possible situation. When we find that our character conflicts with our values, we can set out to revise that character and engage in self-correction. The patterns of acting and desiring that constitute our way of life do not force us to follow them. On the basis of such considerations, Michele Moody-Adams (1990), for example, concludes that we have an ability to act out of character, and this ability is the reason that character is not destiny. Alternatively, we can note

that traits of character do not fully determine actions—traits may be only probabilistic predictors of behavior.

Such arguments, however, do little to provide a rationale for appealing to the will. Granting that we can sometimes act “out of character”, or resist our habitual ways of acting, raises the question of just what sorts of things can incline us away from those habitual ways. Say that a typically courageous person fails to act courageously on a given occasion. Is this evidence that courage does not in fact determine her actions? It is reasonable to think that, if the agent really had a robust disposition to act courageously, some abnormal features were operating here. Perhaps she encountered a situation all the courageous habituation of her past hadn’t prepared her for; maybe she was simply fatigued or overly emotional (Goldie 2004a); or maybe another disposition provided a weighty enough consideration. Baier illustrates that “[w]hen we act out of character in some respect, it will be still in character in another respect” through a series of examples drawn from Hume’s *History*; “[e]ven Becket, when he donned the character of sanctity and started mortifying his hitherto well-indulged flesh, still showed his old concern for how he looked—sackcloth peeping out from his archbishop’s robes” (Baier 2009, 254). While people do act in unpredictable ways, this shows at best that we did not know enough about them or their circumstances to make confident predictions; that features of their character combined with features of their situation to produce an unexpected result.

Nor does the claim that traits of character do not fully determine actions help much. The word “determine” is often used in a strong sense that does not allow for exceptions, and Moody-Adams’s argument may be drawing on this sort of usage. It may well be false that character determines agency in the sense that if we know everything about a person’s character and everything about the environmental factors she is exposed to, we may still be unable to predict her behavior with absolute certainty. However, as just noted,

whatever determined the action had to come from somewhere—from the agent’s motivational structure or from the environment. If something about the action was not so determined, this can at best show us that determinism is not complete. In this case, we can rephrase from the claim that character determines agency, to the more nuanced one that insofar as agency is determined, it is determined by character.

But there is no obvious opening for the will here; if there is any role for it to play, it is as either a feature of character or as an overdeterminant of features of action already determined by character. Rather, the argument seems to be based on the mistaken view that character consists simply of neatly defined traits of character. If we recognize that character involves far more complexity, however, it will quickly start to look as if “acting out of character” just involves acting on less dominant—or heretofore less noticed—aspects of one’s character. Alternatively, if we want to hold on to the idea that character is constituted entirely or primarily by character traits, we can adopt a context-relative view of such traits to allow that, for example, being honest when threatened with physical force and being honest when one can profit by lying are distinct traits (Upton 2005). Features of one’s character can easily be at odds. We can see this by asking why the agent decided to act out of character. It will be hard to answer that question without appealing to other dispositions to act. Arguments of this sort play a common and familiar role in undermining free will and moral responsibility.

## **II. Forms of Eliminativism**

I have been arguing that it is difficult to avoid eliminativism, and this is borne out in the literature. In the 19<sup>th</sup> century, for example, the thought that character, together with environmental cues, determines action seems to have been commonplace, and frequently came bundled with a wider kind of determinism. Schopenhauer, for example, built a complex metaphysical system on which each individual’s behavior over the course of a life is fully

fixed by their phenomenal character, itself a manifestation of an underlying noumenal character, and to illustrate his point he compared what seems to us self-determined, chosen behavior with the development of an ice crystal—the crystal may appear unique, but it develops according to the necessary rules of its crystalline nature (Schopenhauer 2010, 205). At the (otherwise) opposite end of the philosophical spectrum, Mill insisted that “given the motives which are present to an individual’s mind, and given likewise the character and disposition of the individual, the manner in which he will act might be unerringly inferred” (Mill 1988, 23). But unlike Schopenhauer, he largely dispensed with metaphysics, noting only that “this proposition I take to be a mere interpretation of universal experience, a statement in words of what every one is internally convinced of” (idem).

In the twentieth century the strategy is not significantly different. Powerful compatibilist approaches to the free will debate have built upon Harry Frankfurt’s work on the idea of identification: I act of my own free will, on this view, when my second order volition happens to conform to the desire I actually act on (Frankfurt 1982); or, alternatively, in later work, when I am satisfied with the desires I act on (Frankfurt 1992). But on this view what I do is either not up to me at all—if, for example, I am in the grip of forces I do not identify with but cannot defeat—or it is a result of the sort of character I happen to have. A similar theme runs through the work of Bernard Williams, who views character as constituted by our projects and desires, so that we can make no decisions unless they are supported by those desires (Williams 1993; Williams 1981a). This tendency is developed further by authors like John Fischer, who see intentional actions as actions stemming from moderately reasons-responsive mechanisms, which we form and take responsibility for in the course of normal development (Fischer and Ravizza 1998).



It might seem that I am being unfair. After all, the authors just mentioned are specifically attempting to explain such notions as the will, free will, and decision, among others. But in their attempts to do so, we find a clear leaning toward explaining these concepts in terms of, and as dependant on, a prior and more robust notion of character. Sometimes the attempt is implicit; other times it is explicit, as in Williams's famous argument that all reasons are internal reasons—that nothing can count as a reason for me to act unless there is a clear path to that reason from my subjective motivational set which, of course, Williams tends to identify with my character (Williams 1981b; Williams 1973). Even when the move is implicit, however, my point is that such accounts seem to make a separate volitional capacity redundant; character can do all the explanatory work.<sup>2</sup>

There are similar strategies on the other side. The logic is, again, simple. Whenever we encounter a choice, we are capable of exercising our capacity for willing by making a decision about how to act. True, our character might incline us in one direction or another, but character is not deterministic; we can choose which inclination to follow, and we can do so because we have a will. If we go down this path, however, the role of character retreats. Character does not determine action at all; rather, it suggests courses of action, which we validate or invalidate by our will. This threatens, however, to reduce character to little more than the Summary view, on which character traits simply are patterns of behavior. This view has been roundly criticized, among other reasons, for being unable to make sense of the explanatory role appeals to character can serve on such a conception (Brandt 1970; Gilbert 2006). The explanatory work is being done entirely by our acts of willing in particular situations; if any patterns emerge in our willing, those patterns have no existence independent of the instances of willing

---

<sup>2</sup> For a detailed account of Williams on character and its relation to an agent's subjective motivational set, as well as a critique of his attempt to reduce the will to character, see Altshuler (2013).

themselves. Volitional explanation simply displaces character explanation. The Summary view also allows that character traits serve a predictive function, but cannot explain why. If the traits just *are* patterns of past behavior, there is no reason to think that they can have any bearing on future behavior unless there is a further mechanism at work motivating the past actions and (thereby) making future actions along similar lines more likely. The view thus fails to make sense of the function of character, reducing all agency to willing.

Now, such an approach may appear too extreme, if for no other reason than because eliminating character altogether seems to do too much violence to our ordinary ways of understanding and evaluating agents. The alternative is to find a more palatable view, one on which our actions are indeed determined by character, but that character is the result of willing. Robert Kane is perhaps the most notable proponent of such an account. Arguing that we could not be responsible for our actions unless we chose them, Kane maintains that while our actions are determined by character in most cases, in rare instances we are capable of Self-Forming Willings, that is, volitions that bring character traits into existence (Kane 1996). By freely choosing to lie in one situation, for example, I do not merely choose that one action, but open myself to a future of being open to dishonesty. This view seems to preserve a role for character—in fact, it stresses that the majority of our actions are entirely fixed by character—but the appearance is deceptive. Character now becomes a mere diachronic effect of willing. More significantly, character is always fully open to revision: if I can will new character traits into existence, there is in theory, if not in practice, no limit to what I can will.

Again, my point is not that proponents of character are necessarily eliminativists about the will or vice versa. And there are, in fact, some ingenious attempts to reconcile the two sides. Developing Sartre's account,

for example, Jonathan Webber argues that character may be fixed by projects, which may or may not be compatible with a fairly high degree of freedom of choice (Webber 2006). This is reasonable; it can hold that character inclines us in certain directions, but the will is needed to take the extra push to action. But any such view, which attempts to separate character and will into distinct phenomena, is likely to be unstable, for reasons that I have been pressing. In fact, I suspect that the best strategy is to recognize the common nature of character and will. The question is whether such a strategy must take an eliminativist form.

### **III. Distinct Roles**

Before moving on to that question, we should examine whether eliminativism can be carried out, in either direction, without serious loss. There are good reasons for thinking that character and will are distinct phenomena: they seem to serve distinct functional roles. Character, among other purposes, serves the function of providing third-personal explanations of behavior; will, again among other functions, aims at providing justification. That I have a jealous disposition, for example, may go a long way toward explaining why I have planted a GPS tracker on my partner's person; it does nothing to justify it. The fact that I *think* I have reason to plant the tracker is itself explicable in terms of my character traits, moreover, and thus it may seem as if there is little for my will to do—between my character and my action, will can serve at best an intermediate position; in acting as I do, I am merely playing out my character, much the way Schopenhauer's ice crystal merely follows the directives of its crystalline structure. Suppose, for example, that I were capable of refusing to plant the tracker—of returning it to the store and never thinking of it again. If I could do that, perhaps some other feature of my character can explain it; say, the tension between trust and jealousy, and the importance to me of trust, has gotten to me. But if this is *not* the explanation, if I simply *chose* to return the tracker and that is all

there is to the story, it is not clear what it would mean to describe me as a jealous person, to attribute a character trait to me.

Let's say that we settle for the first option: that I return the tracker because some disposition of my character, rather than my will, outweighed my jealousy. On a popular strain of thought about the will, it is in fact my character that sets my goals; all that's left is to act them out. This is why Aristotle stresses that deliberation is about means, not ends; it is virtue that fixes the appropriate ends. Why, on such a view, would justification be missing? Why isn't "doing this would be incompatible with trust" a good enough justification? The problem is that the justification wouldn't, in this case, provide an explanation of my action. I did not in fact return the tracker because trust gave me reason to do so, but because trust happened to be a trait of my character. So while this picture is perfectly compatible with my acting well or badly, in ways that can and those that cannot be justified, it is not compatible with my having a justification of my action that can explain why I so acted.

The role played by the will in allowing for justification of my action is a normative one. The role of the will on this picture is not merely to allow me to act well, but to allow me to *choose* to act well. What is necessary is that the will be able to serve an executive function. We can see this more clearly when we think about acting badly, especially when we do so in the knowledge that we are doing so—that is, in cases of *akrasia*. Say that I recognize that my planting the tracker is incompatible with trust, and that I ought to do what is compatible with trust. If I set it anyway, it seems I have acted wrongly. But unless I chose to do so, it is unclear that I have really *acted*, and by extension that my "action" can have any moral status. That is, if my action stemmed from a defect in my character's ability to pursue the good, it is unclear that it was an action at all, rather than something more akin to a spasm. This is what seems to follow on the view that *akrasia* involves an overriding of the will by

passion. Of course here I am assuming that choice—and thus will—is necessary to a bit of behavior's being an action. On the character side, we can reject this assumption. We might stipulate that anything brought about by any aspect of my character, broadly construed, counts as an action. Or we might hold that actions, to count as *actions*, i.e., doings attributable to the agent, must stem from fixed dispositions, or perhaps from dispositions that are representative of the agent in some sense that can be further specified. In these cases, counter-normative agency will qualify as genuine agency, but it will be unclear how we might hold someone responsible for it or speak of being obligated to avoid it, and attribution to an agent along such lines threatens to sever the connection with authorship. The exercise of counter-normative agency is thus frequently cited as requiring an account of the will, since without it *akrasia* becomes mysterious as an exercise of agency, or conversely threatens to sever agency from normativity and authorship altogether (Watson 2003; Wallace 2006).

This latter point ties in with some standard arguments about responsibility, since it seems as if to be *responsible* for our actions, we must be capable of actively choosing them on the basis of reasons. Actions that merely stem from character, it seems, can be responsible ones only on the assumption that we choose our characters. This view is common to positions as diverse as those of Robert Kane and Galen Strawson; Kane (1996) argues that we can be responsible in acting from character only insofar as our character stems from prior self-forming actions, while Strawson (1994) holds that responsibility is impossible, since all such actions must themselves be determined by character. Yet both take character without will to be a bare given, something for which no one can be responsible.<sup>3</sup> At the executive level, it seems we are able to act out of character—that is precisely how people can turn their lives around, by choosing to act counter to their dispositions. But

---

<sup>3</sup> For a position similar to Strawson's, see Chapter VII in Nagel (1986).

once again, we are faced with a problem: if we *can* act counter to our dispositions, this seems to suggest that they are only dispositions to act insofar as we endorse them or choose to act on them. And if that is the case, it is not clear that they really are dispositions to start with. Moreover, leaving the will out once again seems to cut us off from authorship of our character; that character may change, but with no will driving the change, it is a blind process without an author.

It looks like we cannot simply eliminate the will in favor of character, because we will then lose the executive power that comes with it, and which is needed for justification, responsibility, and control. But we cannot simply settle for the opposite view, that of eliminating character in favor of the will. For if we do that, we will leave it utterly mysterious how it is possible to predict the future actions of ourselves and others, or just why it is that we seem to make similar decisions over time—that is, why our actions come in patterns. As I've noted above, attempting to eliminate character in this way will leave us with nothing more than the Summary view—the view that character, if there is such a thing, is just the sum total of our intentional actions. But if each exercise of agency in this scenario is caused by an individual act of willing, dispositions of character serve no *explanatory* role. What, then, would license me to infer anything about your future behavior from your past behavior?

There is a further problem in the wings. Let's say that we accept an account on which the will has primacy and character is threatened by a corresponding reduction to the Summary view. We will lose more than just intuitiveness. For what will motivate us to make choices or commit to following through on them if not character? The executive power by which we commit ourselves to our resolutions seems to presuppose our having the dispositions to care about and bind ourselves to those resolutions. This seems to me to be one of the major weaknesses of many will-centered

accounts, such as that presented by Korsgaard (1996; 2009). What unifies the agent's actions with cross-temporal principles, on this view, is just something like the agent's practical identity. But so long as that identity is itself constituted by nothing other than practical commitments, it is far too thin to make sense of agents' actually sticking to their identity in the face of conflict rather than abandoning it at the first sign of trouble. Thus, responding to Michael Bratman's view of planning for the future as structured by higher order policies (Bratman 2000), Goldie notes that "people can, and often do, lead unreflective lives, being playful, patient, temperate, forgiving, modest, and so on, without giving any reflective thought to their policies" (Goldie 2012, 96). My point here is stronger: it is not simply that we typically do not need to think about our policies insofar as they are encoded into our dispositions and do not require explicit thought to adjust to, but that policies or commitments cannot maintain their intra-personal diachronic authority without support from dispositions. Aside from having principles and seeing that they have practical reason to abide by those principles, agents also need to *care* about the principles, at least implicitly.

This point is the main appeal behind a rival account such as that developed by Frankfurt (2006), on which what binds us to a particular identity is that we *care* about it, and caring involves wanting to go on caring. As I've already suggested, Frankfurt's account takes things to the opposite extreme of eliminating the will or at least reducing it to character. We can attempt to split the difference with a beefed up account of the will, on which it is not simply a capacity for choosing, but also for creating new motivations, and something like this view seems to me to be right (Davenport 2007). But such motivations must still link up with a preexisting motivational framework. And separating this power to create new motivations from any dispositional background leaves any exercise of the power itself unmotivated.

#### IV. The Single Phenomenon View

If character and will serve distinct functions in this way, it seems that neither can be eliminated in favor of the other. But the functional distinctions I've just discussed are more fluid than I've made them appear. If we can establish what counts as a reason, or a good reason, independently of willing,<sup>4</sup> then responding to reasons (or to good reasons) in certain circumstances may itself be a character trait.<sup>5</sup> This would at the very least give agents the power to act in accordance with—or contrary to—right reason without requiring recourse to a specialized faculty of willing, though admittedly we would be left with no control over our exercise of this power of responding to reasons itself. Reasons-responsive views pay the price of losing the sense that it is the *agent*, not merely her dispositions or mechanisms, that responds to reasons. It is in part for this reason that Fischer and Ravizza, for example, complement their view with the agent's *ownership* of her reasons-responsive mechanisms, though it is unclear that this addendum can succeed. If those mechanisms can fail to belong to the agent, it seems, so can the mechanisms by which agents come to own them (Shabo 2005); unless, perhaps, the agent's distinctive activity as will comes into play. Yet such views do blur the boundaries between will and character by bestowing some of the former's normative function on the latter. If, on top of this, we recognize that our decisions are constitutive of character (and vice versa), the case for thinking a distinct

---

<sup>4</sup> Some versions of Kantian constructivism, for example, follow just the line I am leaving out here. Korsgaard (2008), for one, develops the view that for something to count as a reason, it must be willed, or at least must be capable of being willed by an ideal agent (her view is a bit sketchy on the details). But even on this view, it is not the case that just *anything* could count as a reason, since there are things ideal agents must will, and things they cannot will. If so, we can treat constructivism—for the purposes of this paper—as a type of moral realism, on which reasons are established independently of any subjective exercise of willing.

<sup>5</sup> Something like this view is suggested by McDowell (1998). Goldie argues that “all character traits are *reason-responsive*” (2004b, 13), though he is working with a much narrower version of character than one I have been discussing, since on his view character refers to deep ethical dispositions and is distinguished from the sorts of non-moral mannerisms and ways of behaving he places under the heading of personality.



capacity is necessary for executive functions is weakened: character will itself be an executive faculty, by virtue of responding to reasons and giving rise to (and itself embodying) decisions.

So instead of separating character from will as two distinct phenomena, we may do better to think of our agency as having a dual nature, neither fully passive nor fully active, a view that has the added benefit of being able to explain just *why* we can both explain (in non-rational terms) and justify the very same actions without requiring that giving two such explanations necessarily overdetermines the action. The space of reasons need not be ontologically *sui generis*. But this dual nature should not give us reason for eliminativism. Without recourse to character, we would lose our strongest explanations of why certain agents are likely to *consistently* respond to relevantly similar reasons in similar ways; and we would lose also the ability to explain what motivates even justified action. But without will, while we could establish more and less desirable ways of acting, we would lose the connection between action on one hand, and normativity and responsibility on the other.

I have been assuming throughout that reduction of character to will, and vice versa, is roughly equivalent to eliminativism. This suggestion can be fleshed out through an analogy: suppose that one argues that (1) all mental phenomena just are, in reality, physical phenomena. This seems like a straightforward eliminativism: all that is left, on such a view, is to figure out the exact physical specifications of each mental phenomenon and dispense with the folk terminology. But one could also maintain that (2) although mental phenomena just are physical phenomena, we may retain the folk terminology because we do not yet know how to match it fully to the physical vocabulary, although with the caveat that even if the folk terminology is too slippery to simply replace by physical terminology, ultimately it refers—though in a vague way—only to fully physical phenomena. This, too, seems to

be a kind of eliminativism, since the distinction between the mental and the physical is a merely semantic one. Finally, one can hold that (3) although mental entities exist, they are entirely constituted by physical ones. This, I suppose, is what is normally thought of as reductionism, though sometimes the term is used for (1) or (2). But I am not convinced that it avoids being substantially different from eliminativism: unless we are willing to grant that mental phenomena have functions that physical phenomena themselves do not (for example, to repeat an old argument, mental phenomena are intentional, and it is not clear how physical phenomena could be), we continue talking about mental phenomena only provisionally—perhaps we do not know *how* to say what we need to say in purely physical vocabulary, but the conviction is that it must be possible to do so, if perhaps not at our present stage of scientific development.

I do not here wish to maintain that reductionism is *necessarily* a form of eliminativism. To say that smoke is a collection of airborne particles, or a toupee a contraption of artificial hairs held together by cloth and glue, surely reduces the smoke and toupee without eliminating them; the smoke is real, as is the toupee, in that they really are their constituents arranged in a certain way. Nor do I want here to endorse the claim suggested in the above example, that reductionism about the mental just is eliminativism about it. But it seems to me that in reducing character to will or vice versa, one is generally committed to an eliminativism. Of course, following strategy (2) or (3), someone like Williams can continue to use volitional terms, despite holding that there is nothing behind them but character. But in that case either the usage is ultimately obfuscating an underlying disconnect, or it simply helps itself to volitional features it cannot in fact grasp.<sup>6</sup> So if I am

---

<sup>6</sup> For a justification of this claim in Williams's case, and an argument for the claim that it is insufficient to explain the volitional features it relies on, I again refer readers to my longer discussion in Altshuler (2013). A quick example: Williams argues that "if one acknowledges responsibility for anything, one must acknowledge responsibility for decisions and action which are expressions of character" on the grounds that expressing an agent's character is paradigmatic

right and character and will really do have distinct functions, then one cannot be reduced to the other without eliminating it. As I suggested above, *some* functions of the will may be performed by character, and the reverse is likely true as well. But complete reduction of one to the other will leave a functional remainder (this is one reason that the point does not apply to the toupee or smoke examples above: neither a toupee nor smoke has any function that cannot be equally well filled by its constituents). This gives us reason to think that character and will must be kept discursively separate in order to accommodate their functional distinctions, but not to claim that they must be ontologically separate as well.

Isn't that last concession a kind of eliminativism? It might seem so, if we think the suggestion is that there is a specific kind of phenomenon to which I propose ascribing the functions of both will and character. So, for example, if the claim were that our agency can be completely described using only physiological terms, and that this capacity then expresses itself in diverse functions that can be roughly described in character terms or volitional terms, the proposal would indeed be eliminativist. But I am not claiming that agency is any completely describable thing. No doubt there is a physical vocabulary in terms of which our agency can be partially described, and no doubt such a vocabulary serves important functional purposes. But it does not follow that any one description is *complete* and thus capable of serving as a ground to which all other descriptions can be reduced. As agents, we do things that are frequently predictable, that are legitimate targets of praise and blame, that express patterns of activity over time, that respond to reason and deliberation and, yes, that involve neural firings and muscle fibers. We need diverse vocabularies to grasp these different functions; no one of them is sufficient.

---

of being an agent's own (Williams 1981a, 130). But it is not clear how one can explain *why* being one's own brings responsibility with it if ownership rests on being a feature of the agent's subjective motivational set rather than choice.

Of course this is not to deny that a vocabulary *could* be devised that can capture all of the functions of agency. My claim here is simply that the vocabulary of character cannot replace that of will, nor vice versa, and that we have no reason to think the two are fully reducible to any existing third vocabulary, even if we have reason to think that they are not ontologically distinct determinants of agency. Davidson famously argued that mental and physical events are the same events under different descriptions. Here I am making a similar claim: acting from character and willing an action can give us two different descriptions of the same exercise of agency. Some descriptions will allow us to draw conclusions about how the agent is likely to act in the future; other descriptions will allow us to hold the agent responsible. But my view also differs in important ways from Davidson's model. First, while Davidson of course did not hold that every physical event can also be described as a mental event, he did hold the reverse. But I make no such claim. Some of our actions from character are ones we could not, without significant distortion, describe as volitional exercises. But it is also true that some choices are ones we would be hard pressed to attribute to character, except in the vaguest terms—by noting that an agent has a disposition toward arbitrary behavior, for example. This is one reason the use of distinct vocabularies is often called for.

But many exercises of agency admit equally well of both kinds of descriptions, and here is the second difference from the Davidsonian model: while it is often appropriate to describe the same exercise of agency in both kinds of terms, the descriptions can easily blend together. Thus, we may say of the inveterate procrastinator that he has chosen to put off yet another task, and we may likewise, as noted above, speak of normative features of character. The reason such cross-talk is possible is that character and will do not belong to sharply delineated spaces, the way that physical and mental descriptions supposedly belong, respectively, to the space of law and the space of reasons. The latter distinction is established, in my view, less

because there is a sharp divide in nature to which it corresponds and more as a bulwark against the creeping fingers of reductionism. In the case of character and will, their cross-bleeding and nature as determinants of action gives rise to the same reductionist tendencies. But such reductionism, I have been urging, should be resisted without a corresponding insistence on a division in nature. If the aim of reductionism is simplicity, we can satisfy that aim by drawing sharper lines between various normative and explanatory functions of agency and grouping those functions into families of descriptions with less distortion than obtains when we attempt to eliminate the boundaries altogether and either reduce all acting on reasons to features of agents or reduce all constancy to a mysterious repetition of free acts.

The distinct functions of will and character, in other words, need not require us to take will and character as two distinct phenomena, or two ontologically distinct capacities, and we can thereby address the problem of overdetermination. But the distinct functions should, at the very least, require us to adopt a dualism in our talk about agency. Understanding agency should not demand that we attempt to explain character in terms of the will or the will in terms of character. But it should lead us to see agency as having two distinct functional roles, with the implication that a proper understanding of the relation of character to the will must lead to a careful delineation of the functions of both so as to allow for our practices of both explanatory and normative discourse, although there may well be a great deal of terminological and functional ambiguity at the edges. An explanation of how we can choose on the basis of motives without thereby undercutting the normative authority of choice demands ontological parsimony along with discursive multiplicity.

## References

- Altshuler, Roman. 2013. "Practical Necessity and the Constitution of Character." In *The Moral Philosophy of Bernard Williams*, edited by C. D. Herrera and Alexandra Perry, 40–53. Newcastle: Cambridge Scholars Publishing.
- Baier, Annette. 2009. "Acting in Character." In *New Essays on the Explanation of Action*, edited by Constantine Sandis, 241–256. Basingstoke: Palgrave Macmillan.
- Brandt, R. B. 1970. "Traits of Character: A Conceptual Analysis." *American Philosophical Quarterly* 7 (1): 23–37.
- Bratman, Michael E. 2000. "Reflection, Planning, and Temporally Extended Agency." *The Philosophical Review* 109 (1) (January): 35–61.
- Butler, D. 1988. "Character Traits in Explanation." *Philosophy and Phenomenological Research* 49 (2): 215–238.
- Davenport, John J. 2007. *Will As Commitment and Resolve: An Existential Account of Creativity, Love, Virtue, and Happiness*. Fordham University Press.
- Fischer, John Martin, and Mark Ravizza. 1998. *Responsibility and Control*. New York: Cambridge University Press.
- Frankfurt, Harry. 1982. "Freedom of the Will and the Concept of A Person." In *Free Will*. New York: Oxford University Press.
- . 1992. "The Faintest Passion." *Proceedings and Addresses of the American Philosophical Association* 66: 5–16.
- . 2006. *Taking Ourselves Seriously & Getting It Right*. Stanford: Stanford University Press.
- Gilbert, Margaret. 2006. "Character, Essence, Action: Considerations on Character Traits after Sartre." *The Pluralist* 1: 40–52.
- Goldie, Peter. 2004a. "What People Will Do: Personality and Prediction." *The Richmond Journal of Philosophy* (7): 11–18.

- . 2004b. *On Personality*. Routledge.
- . 2012. *The Mess Inside: Narrative, Emotion, and the Mind*. Oxford: Oxford University Press.
- Hieronymi, Pamela. 2009. "The Will as Reason." *Philosophical Perspectives* 23 (1): 201–220.
- Hudson, S. D. 1980. "Character Traits and Desires." *Ethics* 90 (4): 539–549.
- Kane, Robert. 1996. *The Significance of Free Will*. New York: Oxford University Press.
- Korsgaard, Christine. 1996. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- . 2008. *The Constitution of Agency*. New York: Oxford University Press.
- . 2009. *Self-Constitution: Agency, Identity, and Integrity*. New York: Oxford University Press.
- McDowell, John. 1998. "Virtue and Reason." In *Mind, Value, and Reality*, 50–73. Cambridge, MA: Harvard University Press.
- Mill, John Stuart. 1988. *The Logic of the Moral Sciences*. Chicago: Open Court.
- Moody-Adams, Michele. 1990. "On the Old Saw That Character Is Destiny." In *Identity, Character, and Morality: Essays in Moral Psychology*, 111–131. Cambridge: MIT Press.
- Nagel, Thomas. 1986. *The View from Nowhere*. New York: Oxford University Press.
- Schopenhauer, Arthur. 2010. *Schopenhauer: The World as Will and Representation*: Translated by Christopher Janaway. Cambridge University Press.
- Shabo, Seth. 2005. "Fischer and Ravizza on History and Ownership." *Philosophical Explorations* 8 (2): 103–114.
- Smith, Michael. 2012. "Four Objections to the Standard Story of Action (and Four Replies)." *Philosophical Issues* 22 (1): 387–401.
- Strawson, Galen. 1994. "The Impossibility of Moral Responsibility." *Philosophical Studies* 75: 5–24.

- Upton, Candace L. 2005. "A Contextual Account of Character Traits." *Philosophical Studies* 122 (2): 133–151. doi:10.1007/s11098-004-6218-6.
- Velleman, J. David. 1992. "What Happens When Someone Acts?" *Mind* 101: 461–481.
- Wallace, R. Jay. 2006. *Normativity and the Will*. New York: Oxford University Press.
- Watson, Gary. 2003. "The Work of the Will." In *Weakness of Will and Practical Irrationality*, edited by Sarah Stroud and Christine Tappolet, 172–200. New York: Oxford University Press.
- Webber, Jonathan. 2006. "Sartre's Theory of Character." *European Journal of Philosophy* 14: 94–116.
- Williams, Bernard. 1973. "The Makropulos Case: Reflections on the Tedium of Immortality." In *Problems of the Self*, 82–100. New York: Cambridge University Press.
- . 1981a. "Practical Necessity." In *Moral Luck: Philosophical Papers, 1973-1980*, 124–31.
- . 1981b. "Internal and External Reasons." In *Moral Luck: Philosophical Papers, 1973-1980*, 101–113. New York: Cambridge University Press.
- . 1993. "Moral Incapacity." *Proceedings of the Aristotelian Society* 93: 59–70.