




# The essential superficiality of the voluntary and the moralization of psychology

Matthieu Queloz<sup>1</sup> 

Accepted: 12 August 2021  
© The Author(s) 2021

**Abstract** Is the idea of the voluntary important? Those who think so tend to regard it as an idea that can be metaphysically deepened through a theory about voluntary action, while those who think it a superficial idea that cannot coherently be deepened tend to neglect it as unimportant. Parting company with both camps, I argue that the idea of the voluntary is at once important and superficial—it is an essentially superficial notion that performs important functions, but can only perform them if we refrain from deepening it. After elaborating the contrast between superficial and deepened ideas of the voluntary, I identify the important functions that the superficial idea performs in relation to demands for fairness and freedom. I then suggest that theories trying to deepen the idea exemplify a problematic moralization of psychology—they warp psychological ideas to ensure that moral demands can be met. I offer a three-tier model of the problematic dynamics this creates, and show why the pressure to deepen the idea should be resisted. On this basis, I take stock of what an idea of the voluntary worth having should look like, and what residual tensions with moral ideas this leaves us with.

**Keywords** Voluntariness · Bernard Williams · Conceptual ethics · Moralization of psychology · Moral luck · Responsibility

---

✉ Matthieu Queloz  
matthieu.queloz@wolfson.ox.ac.uk

<sup>1</sup> Wolfson College, University of Oxford, Linton Road, Oxford OX2 6UD, UK

## 1 Introduction

What makes an action voluntary? And does it matter? Is the concept of the voluntary even an important idea to have in our conceptual repertoire? Theorists who have thought the concept of the voluntary important have tended to do so because they considered it an enigmatic but profound idea that could be deepened through a theory of action, while those who thought it a superficial idea that could not coherently be deepened have tended to neglect it as unimportant.

Parting company with both camps, I contend that the idea of the voluntary is at once important and superficial—indeed, that it is important only as long as it remains superficial. The notion of the voluntary is an *essentially* superficial notion that performs important functions, but it can only perform those functions if we refrain from deepening it.

Something like this view might be thought to be adumbrated in G. E. M. Anscombe's (2005, 2008a) essays from the 1960s as well as in certain passages of P. F. Strawson's (2008) "Freedom and Resentment,"<sup>1</sup> but it is most explicitly articulated in the following passage from Bernard Williams's *Shame and Necessity*:

[T]he idea of the voluntary ... is essentially superficial. It is a mistake to suppose that the notion of the voluntary is a profound conception that is threatened only by some opposing and profound theory about the universe (in particular, to the effect that determinism is true). That supposition underlies the traditional metaphysical problem of the freedom of the will. ... Just as there is a "problem of evil" only for those who expect the world to be good, there is a problem of free will only for those who think that the notion of the voluntary can be metaphysically deepened. In truth, though it may be extended or contracted in various ways, it can hardly be deepened at all. What threatens it is the attempt to make it profound, and the effect of trying to deepen it is to put it beyond all recognition. (Williams, 1993, pp. 67–68)

Although this passage may make Williams sound like a complacent compatibilist, we shall see towards the end of this paper that this impression is misleading.<sup>2</sup> The question more immediately raised by this passage, however, is what the thesis of the "essential superficiality" of the idea of the voluntary amounts to, and how a "superficial" idea of the voluntary contrasts with a "deepened" one. I believe that Williams is on to something important in this telegraphic passage, and in developing the thesis as I propose to understand it, I shall consult his scattered remarks bearing on

<sup>1</sup> The views on the voluntary and its ethical role that Anscombe develops in those 1960s essays as well as in her second McGivney lecture (2008b) differ notably from her remarks on voluntary action in *Intention* (1957, §§7–17, 20–22, 49). See Bierson and Schwenkler (forthcoming). On the Strawsonian aspects of this Williamsian view, see Russell (2017b, forthcoming), Queloz (2021b), De Mesel (manuscript), and Emilsson (manuscript).

<sup>2</sup> As Paul Russell also concludes after examining Williams's criticism of the "reconcilers" and the "old compatibilism" (Williams 1995c, p. 6): "Whatever final position Williams arrives at, it should not be understood as any form of comfortable or complacent compatibilism" (P. Russell forthcoming, p. 16).

the topic.<sup>3</sup> But I shall have to go substantially beyond Williams in doing so, since he did not develop the claim in detail. Nor was it developed in the direction I envisage by the subsequent literature.<sup>4</sup> The suggestion that the idea of the voluntary is “essentially superficial” is alluded to, but not discussed, in Duff and von Hirsch (1997, p. 103) and Crisp (2017, p. 1), and insofar as it has been unpacked at all, it has been interpreted as a dismissive remark pointing to superficiality as a defect, with the qualification “essentially” meaning only “basically” or “at bottom.” Yeager (2006, ch. 2), for example, takes the superficiality of the idea of the voluntary to be a flaw to be remedied, while Deigh (2008, p. xi) takes it as an encouragement to move away from the idea of the voluntary in our practices of moral appraisal.

By contrast, I take the superficiality of the idea of the voluntary to be neither a flaw nor a reason to rely less on it in our practices of moral appraisal. On the contrary: the idea of the voluntary is very much worth having, and its superficiality is not a defect, but an important feature of it, one that is “essential” in that the concept cannot function properly without it. We not only need the idea of the voluntary, but we also need it to be superficial. Much turns on the issue, moreover, because the idea of the voluntary marks a key point at which our psychological ideas about action link up with our moral ideas about responsibility, and deepening the idea threatens to rob attributions of moral and legal responsibility of their efficacy in helping us live together.

I proceed as follows: I begin by elaborating the contrast between superficial and deepened ideas of the voluntary (Sect. 2) and go on to identify the important functions that the superficial idea performs in relation to concerns for fairness and freedom (Sect. 3). I then suggest that theories trying to deepen the idea are morally motivated and exemplify a problematic moralization of psychology—they warp psychological ideas to ensure that moral demands can be met (Sect. 4). I offer a three-tier model of the problematic dynamics this creates, and show why the pressure to deepen the idea should be resisted (Sect. 5). Finally, I draw out what an idea of the voluntary worth having should look like (Sect. 6), and what residual tensions with some of our moral ideas this leaves us with (Sect. 7).

## 2 Superficial versus deepened ideas of the voluntary

In speaking of “the” idea of the voluntary, I do not mean to deny that there are several different, though related, concepts that we express with the word “voluntary.”<sup>5</sup> The concept that the word expresses in its moral use, for example,

<sup>3</sup> See Williams (1993, p. 67; 1995a, p. 578; 1995b, pp. 127–28; 1995e, pp. 243, 247n5; 1995i, p. 495; 2006b, pp. 124–25).

<sup>4</sup> Though I have benefited from Moore’s (2003, 2006) and Louden’s (2007) reconstructions of Williams’s critique of the morality system as targeting the concept of a “purely voluntary act,” in Moore’s apt phrase. See also Queloz (forthcoming) as well as Krishnan and Queloz (manuscript) for discussions of that aspect of Williams’s critique.

<sup>5</sup> Williams (2006a, p. 98) notes that the exact contours of the concept will vary with the purposes to which it is put. Anscombe’s own use of the word varies over time, and she even considers variations on

may not be exactly co-extensive with the concept it expresses in its legal use; moreover, the boundary between voluntary and involuntary action may be blurred in the word's moral use, allowing for actions that lie halfway between the fully voluntary and the utterly involuntary, but sharp in at least some of its legal uses: in criminal law, for instance, an accusation must issue in a verdict of guilty or not guilty, and this may force the binary classification of actions into either voluntary or involuntary ones, even if some actions are neither clearly one nor clearly the other.<sup>6</sup> A sufficiently close look at “the” idea of the voluntary may thus find that, in different contexts, it differentiates into a collection of related but subtly different ideas. And, as Alfred Mele (2017, p. 137) has noted, different conceptions of the voluntary may be activated on different occasions, depending on what kind of case one is considering.

But my concern here is to start further back, from a point of view that allows us to ask, more generally and with greater detachment from any particular context of application, why we should want to think in terms of *any* kind of idea of the voluntary, and why the kinds of ideas of the voluntary worth having had better not be deepened.

What marks out an idea (or concept, or notion—I use the terms interchangeably here) of the voluntary as having been deepened? The first sign is typically *epistemic*, i.e. to be found in what is or is not regarded as evidence for the applicability of the concept. When a deepened concept of the voluntary is at work, the question whether the concept applies to an action, i.e. whether agent A *voluntarily*  $\phi$ -ed or not, is treated as *not yet settled* by commonly ascertainable evidence such as the most immediate circumstances and most proximate causes of A's  $\phi$ -ing.<sup>7</sup> Finding that A burned down the house intentionally and was not sleepwalking at the time, for example, would then still leave open the question whether A burned down the house *voluntarily*. What *would* settle the question depends on the theory in terms of which the idea of the voluntary is deepened. But what can be said at a theory-neutral level is this: determining whether the concept of the voluntary applies requires one to look *beyond* such superficially available evidence and trace the action to its ultimate source in the deeper recesses of the causal chain from which the action originated, because the *real mark* of voluntariness is to be found further back—be it further back along the chain of physiological and/or mental events that issued in the action, or further back in the agent's life history.

---

Footnote 5 continued

the concept that would be applicable to non-human animals if we spoke of “desire” instead of “will”; see Bierson and Schwenkler (forthcoming, §3).

<sup>6</sup> See Williams (1995i, 2005a, p. 271; 2006a).

<sup>7</sup> Focusing on what someone *treats* the question of voluntariness as settled by allows me to register perspectival differences between the concept someone else uses and the concept I myself use. In particular, it allows me to describe the concept that someone uses *without myself endorsing* the use of that concept; this difference would be lost if I simply wrote “the question of voluntariness is settled by,” as this would not be to register a fact about *someone else's concept* of voluntariness, but a fact about *voluntariness as I conceive it*, drawing on, and thereby revealing something about, the concept of voluntariness I use.

A well-known example of how a deepened concept of the voluntary can betray itself primarily through the epistemology of its application is Benjamin Libet's experiment, in which subjects hooked up to an electroencephalograph were put in an optimal position to deliberate at their leisure about whether and when to flick their wrists. Yet what motivated the experiment was precisely the perception—which betrays a deepened concept of the voluntary—that the fact that subjects carefully deliberated and acted only when they really meant to did not yet settle the question whether they *voluntarily* flicked their wrist. *That* was to be determined by having subjects note the time at which they formed the intention to flick their wrist and comparing it against the neurophysiological processes underpinning their conscious deliberation and action.

Yet these epistemic marks are but symptoms of the fact that a deepened concept of the voluntary has been *metaphysically* deepened, as Williams put it in the above-quoted passage. A concept of the voluntary has been metaphysically deepened if it presents voluntary action as something that can be cleanly carved off from involuntary bodily movements by conceiving of the agent's will as fully extricable from its enmeshment in contingent forces external to it. Such a deepened concept of the voluntary thus presents voluntary action as being (a) *distinct in nature* from involuntary bodily movements and the contingent forces that determine them, and (b) *pure of contingency*, i.e. capable of being, at least some of the time, entirely unconditioned by and fully isolated from those contingent forces, thereby grounding not just a distinction of degree, but a categorical distinction between what is the product of the agent's will and what is the product of forces beyond it.

As I shall use the term, a concept of the voluntary therefore counts as having been “deepened” in virtue of having been *metaphysically* deepened, presenting its object as something *sui generis* and pure of contingency. It is this metaphysical aspect that is the defining feature of deepening, while the epistemic aspect is its characteristic (but not strictly necessary) manifestation. The metaphysical aspect underlies and explains the fact that a deepened concept tends to be epistemically demanding, requiring one to dig deep into the aetiology of an action in order to determine whether or not the concept applies. Notice that, on this definition, deepening does not admit of degrees. While we might conceivably grade a concept's deepening according to how far back in the action's aetiology the true marks of voluntariness are to be sought, a clean and categorical separation of what is purely a product of the agent's will from what is a product of external forces can only be achieved by offering some metaphysical account of voluntary action that allows the agent's will to be, at least some of the time, entirely unconditioned by forces beyond itself. If the agent's will were always, however slightly, under the influence of forces beyond itself, this would introduce an element of contingency or luck—and hence of involuntariness—into all the agent's actions.

One way to metaphysically deepen the concept of the voluntary is to maintain that A's  $\phi$ -ing counts as voluntary if and only if A's  $\phi$ -ing has the right kind of neurophysiological aetiology. What kind of aetiology exactly would this be in the Libetian example we considered above? This is controversial because the interpretation of the experiment is controversial. But a deepened concept might be thought to be operative, for instance, in interpretations that withhold application

of the concept of the voluntary unless the conscious experience of willing *precedes* the neurophysiological process issuing in the action. This conceives of voluntary action as having its ultimate or terminal causal source in something distinct from neurophysiological processes—the will—that exerts control over those processes.<sup>8</sup> This metaphysically deepened concept of the voluntary is what licenses the inference from the observation that neurophysiological activity precedes conscious willing to the conclusion that the action in question is not voluntary. This also illustrates that a deepened concept of the voluntary can primarily manifest itself *negatively*, i.e. through what is treated as foreclosing the application of the concept. People can be very clear about what does *not* count as the right kind of neurophysiological underpinning of voluntary action without being clear about what *would* count as the right kind.

Yet the historically most prominent way in which philosophers from Descartes through Locke, Hume, and Bentham to Mill have deepened the notion of the voluntary is through what has come to be known as the *theory of volitions*.<sup>9</sup> The theory starts out from the thought that what turns a bodily movement into a voluntary action is its tracing back to a special kind of mental cause, namely an *act of will* or a *volition*; but this in itself does not yet necessarily amount to a deepening of the idea of the voluntary, since this much could be granted by more recent action theorists like O’Shaughnessy (1973, 2008a, 2008b) and Hornsby (1980), even though they develop accounts of acting and willing in terms of the notion of *trying to  $\phi$*  that are precisely not threatened by the possibility that determinism might be true and do precisely not require us “to look back beyond the trying” (Hornsby, 1980, p. 59). What marks out someone like Descartes as operating with a deepened idea of the voluntary is rather the further fact that the act of will or volition is itself understood as something that is distinctively *pure of contingency*—something that lies beyond the reach of empirical determination by powers external to the acting subject. For Descartes, a volition is an action of the mind or soul; and while the soul can affect matter by affecting the pineal gland (which affects the animal spirits which in turn affect the muscles), the soul is not itself affected by anything beyond itself.<sup>10</sup> On such a theory, voluntary actions have their ultimate source in something that lies deep enough to be completely isolated from the blind play of chance. There is an echo here of the Stoics’ attempts to shelter themselves from upsetting strokes

<sup>8</sup> Another interpretation—which is closer to that favoured by Libet himself—is to conceive of the will as the kind of thing capable of *blocking* or *vetoing* the neurophysiological processes resulting in action. That way of securing the agent’s full control over actions traces back to the Stoics, who thought of the *hegemonikon*, the rational and leading faculty of the soul, as a gatekeeper whose assent was necessary for impressions to issue in action.

<sup>9</sup> For a historical overview of the theory of volitions from Descartes through Locke, Hume, and Bentham to Mill, see Hyman (2011, 2015, pp. 1–24). See also Wilson and Shpall (2012), Candlish and Damnjanovic (2013), D’Oro and Sandis (2013), and Glock (2014).

<sup>10</sup> See Descartes (1964–1976, III 372, XI 342). See also Kenny (1972), Alanen (2002), and Jayasekera (2016). As Williams notes, Descartes’s attempt to explain how I can move my body at will in terms of “a kind of internalized psychokinesis” whereby the pineal gland is the only part of the body that is directly responsive to the will has the uninviting consequence that “the only part of my body directly responsive to my will is one which I cannot move at will” (2005a, p. 277).

of fate, and Descartes' theory of the will went hand in hand with an elaborate Neo-Stoic ethic that also treated the emotions as being fully under the will's control (he once urged Elisabeth of Bohemia to look at the bright side of her uncle's decapitation (Schneck, 2019, p. 757n12)). If voluntary action is rooted in something fundamentally distinct from the muck of contingent forces, this categorically distinguishes it from other happenings: it is not just *less* mired in contingency, but, at base, completely pure of it.<sup>11</sup>

Of course, the theory of volitions has by now largely collapsed under the three successive waves of criticism it encountered in the twentieth century.<sup>12</sup> But there are many ways in which the idea of the voluntary might in principle be deepened, not all of which involve postulating some “ghost in the machine” or some influence coming from beyond the material or natural order. One might, for instance, maintain that A's  $\phi$ -ing counts as voluntary if and only if A's  $\phi$ -ing *reflects nothing but character traits and dispositions that A voluntarily acquired*; this also makes voluntary actions radically different from bodily movements that are products of contingent forces, since not only A's  $\phi$ -ing, but even A's *becoming* the kind of agent who is disposed to  $\phi$  in that situation is fully the product and expression of A's will, undistorted by external circumstances. It does, however, presuppose that what character A came to develop was fully under the control of A's will.<sup>13</sup>

But one might also deepen the idea of the voluntary without this implausible presupposition, by maintaining that A's  $\phi$ -ing counts as voluntary if and only if A  $\phi$ -s *exclusively out of motives that are fully responsive to A's deliberation about what motives to have*—a type of account which admits that we do not, originally, acquire our motives voluntarily, but seeks to re-establish the full autonomy of the

<sup>11</sup> See Williams (2005a, p. 271).

<sup>12</sup> James (1981, p. xxvi) and Russell (1921, p. 285) argued that postulating volitions was not required to make sense of action: action arose when the memories of kinaesthetic sensations initially experienced in mere bodily movements were recruited to function as “motive ideas.” Wittgenstein (1958, pp. 151–52) and Ryle then argued that this theory about “occult inner thrusts of actions” had been accepted only “because it had been wrongly supposed that the question, ‘What makes a bodily movement voluntary?’ was a causal question” (2009, p. 54). And finally, Anscombe (1957) and Davidson (1980) influentially emphasized that the best entry-point for action theory was not the idea of the voluntary, but the idea of the *intentional*. See Hornsby (1980, pp. 48–50) and O'Shaughnessy (2008b, pp. 363–84) for rejoinders to Ryle's criticism in particular, and see Alvarez and Hyman (2019) for a historical overview.

<sup>13</sup> The notion of *control* has itself been theorized in various ways. If A's control over her  $\phi$ -ing is understood in terms of the  $\phi$ -ing being produced by some mechanism that is (a) the agent's *own* and (b) *responsive to reasons*, as in Fischer (1994), that notion of control need not be put in the service of deepening the idea of the voluntary (in the carefully articulated compatibilism of Fischer and Ravizza (1998), for instance, it is not). But it does result in such a deepened idea when A is thought only to have control over her  $\phi$ -ing if her  $\phi$ -ing is the product of a mechanism that is, in the most demanding sense, *fully responsive to reasons*, which amounts to a demand for total deliberative control over one's action-producing mechanism. If A's control over her  $\phi$ -ing is understood, rather, in terms of A's being the *causal source* of the action, then this notion of control can be used to deepen the idea of voluntary action by maintaining that A only has control over her  $\phi$ -ing if she is the *terminal or ultimate causal source* of the  $\phi$ -ing (McKenna and Coates 2021, §1.2); and if A's control over her  $\phi$ -ing is understood in terms of her having a *real choice* between  $\phi$ -ing and alternative courses of action, that notion too can be used to deepen the idea of voluntary action by maintaining that A only has control over her  $\phi$ -ing if her choice to  $\phi$  *objectively determines* an event that had remained objectively undetermined up to the moment of choice (Williams 1995c, pp. 12–14).

will through the idea that an agent's first-order volitions—e.g., A's wanting to  $\phi$ —can fully come under the deliberative control of her second-order volitions—e.g., A's wanting the desire to  $\phi$  to be her will. Here the idea of the voluntary is deepened through a certain picture of human psychology, on which one's rational deliberation can exert total control over one's motives. Though it confines itself to the natural and material order, this is also a way of metaphysically deepening the idea of the voluntary, since it offers an account of the nature of voluntary action that categorically separates it from the influence of contingency. And Williams himself seems to have felt that this kind of account fell within the purview of his critique of metaphysical deepening, for just as he criticized accounts of voluntary action that depended on the will standing outside the material or natural order (2011, pp. 216–17), he also criticized naturalistic accounts insofar as they indulged in what he called the “hopeless fantasy” that one might achieve “total deliberative control of all one's dispositions” (1995h, p. 27).

These different ways of deepening the concept of the voluntary contrast with ideas of the voluntary that remain *superficial*. Consider, for example, a concept on which A's  $\phi$ -ing counts as voluntary already if A both (1)  $\phi$ -ed intentionally and (2) was capable of deliberating normally about whether to  $\phi$ —i.e. not sleepwalking, hypnotized, or in some other state suppressing normal deliberation. This is a superficial concept of the voluntary, since it licenses the ascription of voluntariness merely on the basis of easily observable features of an action and its more immediate or proximal circumstances and causes. Unless the notions of *acting intentionally* or *deliberating normally* have themselves been tendentially theorized to this end, such a concept does not commit its adherents to thinking that there is a metaphysically deep and categorical difference between voluntary actions and involuntary bodily movements. It accepts that whatever is a product of the will is inextricably tied up with what is not.

Although the applicability of a superficial idea of the voluntary remains insensitive to, and therefore tells us little about, the action's deeper aetiology beyond what most proximally caused it, this does not mean that a superficial notion of the voluntary *excludes* there being such ulterior explanations. It leaves room for the thought that A really  $\phi$ -s *voluntarily*, but does so because A has been socialized in a certain way, or is in the grip of an ideology. For the purposes of ideology critique, this compatibility of a superficial notion of the voluntary with ulterior explanations is not a flaw, but a feature—as Etienne de La Boétie's 1576 “Discourse on Voluntary Servitude” already indicated, the real question for ideology critique is often not whether voluntariness is really servitude, but why the servitude really is voluntary.<sup>14</sup> If, through critical reflection on how A came to want to  $\phi$  in the first place, some radical tension emerges, it will not be a tension between  $\phi$ 's claim to being voluntary and the fact that A's wanting to  $\phi$  admits of ulterior explanation, but rather a tension between that explanation and A's understanding of what is really in A's interest. By contrast, the application of a deepened notion of the voluntary really *is* incompatible with many ulterior explanations of how A came

<sup>14</sup> See La Boétie (2016). For an account of ideology critique focused on this question, see Rosen (2013).

to intend to  $\varphi$  in the first place, which is why (as we shall see in Sect. 5) deepened notions of the voluntary are operative in sceptical challenges to free will.<sup>15</sup>

What renders deepened notions of the voluntary attractive, however, is that they promise to give philosophers an objective, practice-independent yardstick by which to take the measure of our practices of responsibility attribution. By first developing a self-standing account of what, deep down, *makes* an action voluntary, we achieve a metaphysically grounded understanding of which actions, if any, we are truly responsible for; and given such an independent understanding of responsibility, we can then take the measure of our actual practices of holding people accountable for their actions: we can determine to what extent *attributions* of responsibility can be grounded in *true* responsibility, and to what extent the human concerns animating those responsibility attributions stand a chance of being satisfied.

In the rest of this paper, I pursue the opposite methodological strategy. Instead of seeking a firm basis in a metaphysically deepened idea of the voluntary before assessing how our practices live up to this independently given idea, I propose to let our understanding of what kind of idea of voluntariness is *worth having* grow out of our understanding of our practices and concerns—in other words, I propose to work *from* the human concerns animating our practices *to* the kind of concept that would best satisfy them. And I conclude that any *viable* version of the concept—i.e. any version capable of serving our concerns effectively—will be superficial, because the idea is essentially superficial: it cannot serve our concerns if it is deepened. In saying that the idea of the voluntary is “essentially” superficial, I am therefore not suggesting that the idea has some discoverable “essence” in light of which it turns out to be superficial—precisely not; the argument runs the other way, from the practical concerns fuelling our interest in the voluntary/involuntary distinction to those versions of the distinction that are worth having.

### 3 The need for the idea of the voluntary

My thesis combines a positive argument to the effect that we need the idea of the voluntary with a negative argument to the effect that this idea should not be deepened. Let us begin with the positive argument. Why would we need the idea of the voluntary?

To answer this question, I first build on—and then move beyond—Williams’s suggestion that the idea of the voluntary is put within reach already by two distinctions that human beings everywhere will find worth drawing: (i) the distinction between acting *intentionally* and acting *unintentionally*; and (ii) the distinction between acting *in a normal state of mind* and acting *in an abnormal state of mind*. As Williams argues, there are practical pressures on individuals living

---

<sup>15</sup> See Pereboom (2001), Waller (2011), and Levy (2011). As Russell (2017a, p. xiv) points out, such sceptical challenges are now taken more seriously than a few decades ago, when the main debate was still between libertarians and compatibilists.

together to be sensitive to both of these distinctions, because their practical necessity follows already from some “universal banalities” (1993, p. 55).<sup>16</sup>

### 3.1 Acting intentionally and in a normal state of mind

First, human beings everywhere need to discriminate between things done intentionally and things done unintentionally. This distinction, applicable not to actions *tout court*, but to actions under a description, is indispensable because it is required to understand an action’s relation to the agent’s practical deliberation, and *what to expect* from an agent whose deliberation bears that relation to this action in such a situation. Williams offers a vivid illustration: when Odysseus and Telemachus confront the suitors of Penelope, they are alarmed to find that although Telemachus hid the suitors’ weapons in a storeroom, the suitors soon distribute those very weapons. Angered by this, Odysseus wonders who opened the storeroom, and Telemachus embarrassedly explains that the mistake was his—that he left the door of the storeroom ajar, but that he did not *mean* to.<sup>17</sup> Telemachus is clearly discriminating here between aspects of what he did that were *intentional* and aspects that were *unintentional*: it was he who left the door ajar, but he did not *mean* to. This shows that even if Homer lacked a direct equivalent of our word “intention,” he had the *concept* of intention—not because we are disposed to draw on this concept in describing the situation, but because Homer and his characters themselves make distinctions which can only be understood in terms of that concept.<sup>18</sup> And it would be surprising if they did not draw *some* distinction along these lines, because sensitivity to which aspects of an action are intentional is crucial to understanding what kind of action it is, and what to expect from one who intends such a thing in such a situation. Had Telemachus *intentionally* left the door ajar, the implication for Odysseus would be alarming, suggesting that Telemachus was not in fact on his side. The ability to discriminate between intentional and unintentional aspects of an action can make the difference between life and death, and it is one that we are bound to have an interest in possessing, because it is crucial to determining the significance of other people’s actions for us. This is the inverse of Anscombe’s (1957, §21) point that the concept of intentional action would not exist if human beings took no interest in each other’s reasons for acting: we can take the fact that human beings everywhere can hardly avoid taking an interest in each other’s reasons for acting as being itself a reason to expect that they will turn out to possess the concept of intention.

Second, human beings everywhere need to discriminate between things people do in what is for them a *normal state of mind* and things they do when they are in what is for them an *abnormal state of mind*. Any interpreter of human action has reason to look not only at which aspects of an action were intentional, but also at the

<sup>16</sup> A methodological approach Williams went on to elaborate in *Truth and Truthfulness* (2002), as I argue in Queloz (2018, 2021a).

<sup>17</sup> See Williams (1993, p. 50). My discussion of Williams’s Homeric examples in this paragraph and the next draws on Queloz (forthcoming).

<sup>18</sup> See Williams (1993, pp. 50–51).

action's relation to the agent's more settled concerns—concerns that, in contrast to momentary whims and impulses, relate to and are possessed for longer stretches of time—and at the extent to which the agent was in a position *to shape the action* to those more settled concerns. If the agent was not in such a position, this renders the action defective in a way that dissociates the action from the agent, since the action fails to express or reflect the agent's more settled concerns. Actions performed in a state of somnambulism or under hypnosis are clearly defective in this sense. But the defect may also lie in the fact that the agent's deliberation was severely skewed or entirely suppressed, as in cases of extreme incident passion. In "Voluntary Acts and Responsible Agents," Williams speaks of such actions as suffering an "inherent or deliberative defect," acknowledging that "[t]he phrase is deliberately wide-ranging, grouping several different kinds of defect, and it is also vague, in shading off into cases in which it is unclear, sometimes for evaluative reasons, whether there is a defect of action" (1995h, p. 24). In other writings on the subject, Williams prefers the terminology of actions done in an abnormal or unusual state of mind. But there, too, he leaves it underdetermined just how capacious the notion of an abnormal state of mind is supposed to be (his examples tend to be the uncontroversial ones of somnambulism, hypnosis, intoxication, delusion, or insanity). For the idea is not that we can specify, in universal and evaluatively neutral terms, which actions are defective or which states of mind are abnormal. Different societies will elaborate the distinction between normal and abnormal states of mind differently, drawing on distinctive concerns and evaluative interests. The very terms in which one makes sense of this distinction depend on what other ideas one brings to bear on the issue. Here also Williams finds a vivid example in Homer: Agamemnon, who seized Briseis from Achilles, did so intentionally, but, as Agamemnon later explains, "Zeus and Fate and Erinys the mist-walking" had "cast fierce *ate*" (1993, p. 52) on his wits, so that he was in a state of delusion or frenzy. By claiming that he was deluded by the gods, Agamemnon is "dissociating the action from himself" (1993, p. 54).

As differently as the distinction will be drawn in different societies, however, the key point is that human communities are bound to draw *some* distinction along these lines, because they are bound to take an interest in whether or not actions stand in regular relations to agents' more settled concerns. Like the distinction between intentional and unintentional aspects of action, some such distinction between normal and abnormal states of mind is also indispensable to knowing *what to expect* from people, because actions performed in an abnormal state of mind are less indicative of how the agent will behave in the future than actions that the agent had a chance to tailor to his or her more settled concerns. Any human community will take an interest in this dimension of action, which is distinct from intentionality. People in abnormal states of mind may still intend to do what they do; but the abnormal state of mind from which the action flows interferes with their capacity to shape their actions to their more settled concerns—either directly, by inhibiting their capacity to act on those concerns, or indirectly, by incapacitating the agents from deliberating properly and preventing these concerns from finding their usual expression in deliberation. Either way, the resulting actions are not representative of more settled concerns.

Now, once these two distinctions are in place, Williams argues, the idea of the voluntary is already within our conceptual reach. He puts this rather strongly at one point by saying that “the idea of the voluntary ... is inherent in the concept of action” (1995e, p. 247n4), but what he must mean is that out of two distinctions that human beings are bound to develop as they try to make sense of each other’s actions, we can already construct the following notion of the voluntary:

A  $\phi$ -s fully voluntarily if  $\phi$ -ing is an intentional aspect of an action that A performs in a normal state of mind.<sup>19</sup>

On this account, “voluntary” action merits the term, i.e. is attributable to the agent’s *voluntas* or will, if and to the extent that it is intended by the agent in a normal state of mind, where neither “intention” nor “normal state of mind” are theorized in a way that deepens the notion of voluntary action. Note that while deepened ideas are dichotomous rather than continuous, categorically separating voluntary action from other bodily movements, this idea of the voluntary is superficial and gradable: the closer an action is to a fully intentional action done in a fully normal state of mind, the more fully voluntary it is. Such an idea of the voluntary enables one to accept what adherents of deepened ideas have to reject, namely that every agent is inextricably enmeshed in a weave of contingent forces in which even the most paradigmatically voluntary action remains, in some respects, in the grip of influences beyond the agent’s will. “One’s history as an agent is a web in which anything that is the product of the will is surrounded and held up and partly formed by things that are not” (1981, p. 29), as Williams puts it. If a fully voluntary action is neither more nor less than an intentional action performed in a normal state of mind, then it is an action that may not have been under the control of the agent’s will *in every respect*, but that was still *as much* under its control as actions ever are, and a great deal more so than an action done unintentionally or in an abnormal state of mind.

But, *pace* Williams, the mere fact that the idea of the voluntary is within the conceptual reach of a set of needful distinctions does not yet show that the idea of the voluntary is itself needed. It merely shows that a set of ideas is needed from which the idea of the voluntary *can* be reached. But having a capacity in principle is one thing; being inclined to realize it in practice is quite another. For example, possession of the concepts *outdoors* and *car* does not by itself guarantee that one will also be an active user of the concept *outcar*, which singles out cars that are outdoors for special treatment—though this possibility is intelligible to us as users of the concepts *outdoors* and *car*, we do not actually think and structure our affairs in those terms. So why did we actually yoke together what is done intentionally and

---

<sup>19</sup> This is my preferred reconstruction of Williams’s notion of the voluntary. The gradability marked by the “fully” comes out in one formula he uses: “an agent does X fully voluntarily if X-ing is an intentional aspect of an action he does, which has no inherent or deliberative defect” (1995h, p. 25); see also Williams (1995h, p. 33n8; 2005b, p. 80n8; 2006a, p. 107). But my reconstruction leans more heavily on the formula he relies on most of the time: “‘A does X voluntarily’ is equivalent to ‘A does X intentionally in a normal state of mind’” (2006b, p. 120). See also Williams (1993, p. 66; 1995f, p. 73; 1995i, p. 495; 2006a, p. 107).

in a normal state of mind under one concept and dignify it with its dedicated linguistic expression? Why did we form a new concept by drawing together just these properties and systematically differentiating in practice between actions that combine them all and actions that do not? Is it really a *new concept*, as opposed to a new word indicating the harnessing of two old concepts to new ends? To answer these questions, we have to understand what *additional concerns* fuel our interest in this particular grouping of properties and lend it its significance.

Even apart from these explanatory gaps in Williams's account, however, there is a more fundamental reason why additional concerns need to be brought into the picture. For the concern that Williams identifies as lying at the root of the concept of the voluntary (more precisely, at the root of its constituent notions) is the concern *to know what to expect from people*. But if this concern were allowed to dictate the contours of the concept of the voluntary unchecked, the resulting voluntary/involuntary distinction would be hard to recognize. In particular, the notion of a normal state of mind would be driven to become a quasi-statistical notion geared towards predictability, and actions would count as voluntary, roughly, to the extent that they were likely to recur. Yet this seems to yield the wrong result, as it is nothing like any of the concepts of the voluntary we know.

Additional concerns therefore have to be factored in if we are to understand why the notion of a normal state of mind as it figures in the voluntary/involuntary distinction seems to be insensitive to mild untypicalities and geared towards ruling out serious mental deviancy. The concepts that offered themselves as building blocks for the concept of the voluntary may basically answer to a concern to know what to expect from people. But in the concept of the voluntary, this concern with predictability is clearly balanced and checked by additional concerns, which pull the resultant concept of the voluntary away from a single-minded focus on what to expect.

### 3.2 The concerns for fairness and freedom

One concern fuelling interest in the idea of the voluntary, I submit, is a concern for fairness in the attribution of responsibility, in the sense in which to carry responsibility for something is to be an appropriate subject of praise or blame with regard to it. Typically, one is treated as being blameworthy or guilty *only if* one acted voluntarily, and the observation that one acted involuntarily *excuses* one from being held responsible in this way (or at least acts as a mitigating circumstance). That arrangement tends to be mirrored in criminal law through what is sometimes called the “voluntary-act requirement.”<sup>20</sup>

<sup>20</sup> See Moore (2010, p. 5) and Saunders (1988). The voluntary-act requirement is related to the *mens rea* doctrine (*actus non facit reum nisi mens sit rea*—the act does not make one guilty unless there is a guilty mind, which is to say a culpable intent; though there are also cases where negligence is enough for criminality). See Hart (1963, p. 40; 2008a, pp. 90–92; 2008b, p. 36; 2008d, p. 14) for a discussion of the connection between the voluntary-act requirement and the *mens rea* doctrine; Duff (2004) argues that the voluntary-act requirement is a further requirement, which he articulates as *mens non facit reum nisi actus sit reus*—the mind does not make one guilty unless there is a guilty act, where “act” is explicitly understood as “voluntary act.”

Of course, certain types of cases call for compromises with more basic concerns, like the need to avoid inherently dangerous situations (e.g. dangerous driving). In those cases, legal responsibility or liability may be “strict,” which is to say independent of the agent’s mental states. Tort law also offers many examples of the voluntary-act requirement being overridden by other considerations: when real damage has been done by something one involuntarily brought about and some response is called for—because *someone* needs to pay, for instance—one might end up being held responsible for something one brought about even if one strained every nerve to keep it from happening.

Generally, however, liability tends to be made conditional on voluntariness, and involuntariness is treated as being exculpatory. The law identifies exculpatory or mitigating circumstances under such headings as “accident,” “mistake,” “provocation” into a passion causing the agent to lose self-control, or “insanity.”<sup>21</sup> The characterization of voluntariness given above fits these excusing conditions. The former two correspond to the requirement that voluntary action be intentional, the latter two to the requirement that voluntary action be done in a normal state of mind. But as H. L. A. Hart notes:

These psychological elements are not *in themselves* crucial although they are important as aspects of responsibility. What is crucial is that those whom we punish should have had, when they acted, the normal capacities, physical and mental, for doing what the law requires and abstaining from what it forbids, and a fair opportunity to exercise these capacities. (Hart, 2008c, p. 152)

The concern for fairness provides a rationale for rendering attributions of responsibility sensitive to the voluntary/involuntary distinction, because our capacity to do the right thing is to a considerable degree hostage to contingent forces outside our control, so that holding people responsible for everything they did would be flagrantly unfair. By focusing responsibility on those actions and aspects of actions that are as much within our control as actions ever are, the voluntary/involuntary distinction mitigates that unfairness. It is, for example, an all too familiar fact that what one ends up having done, once the consequences of one’s intervention have unfolded, is largely a matter of luck, and even something done with the best intentions may issue in deplorable consequences. (Once the flung stone leaves the hand, it belongs to the devil, as a mediaeval proverb has it.) A morality that allocated blame exclusively on the basis of the eventual rather than the intended consequences of actions would turn blameworthiness into a plaything of contingent forces, leaving it largely to sheer luck to decide whether one person attracted more blame than another. That would offend against our sense of fairness by putting the extent to which one attracted blame largely beyond the reach of individual control.

To base moral responsibility notably on the respects in which actions are voluntary, by contrast, is to focus moral appraisal precisely on those aspects of action that are as much as realistically possible within the control of the agent’s will.

---

<sup>21</sup> See Hart (2008b, p. 31). As Hart also observes, much the same conditions are treated as invalidating civil transactions such as wills, contracts, or gifts.

The unintentional aspects of what one does, and the intentional aspects of what one does in an abnormal state of mind, are just the aspects of one's actions that one has less control over. This suggests that it is notably the concern for fairness, conjoined with the empirical fact that the agent's success in realizing his or her intentions depends partly on luck, which drives the focus of responsibility attributions away from the unforeseeable consequences of actions and unusual states of mind and towards the will of the agent as expressed under normal conditions. The voluntary/involuntary distinction helps ensure that moral responsibility is attributed mostly on the fair basis of what is as much as possible within the agent's control.<sup>22</sup>

There is also a second concern that fuels interest in the idea of the voluntary: the concern for freedom as individual self-determination. In order to freely determine the course of their own lives, people need to be able to form determinate expectations as to how the moral code will affect them, and what kind of action they are likely to incur moral sanctions for. If one were subject to blame for everything one brought about, including what one brought about involuntarily, one's blameworthiness would become nearly impossible to anticipate. If one is to retain any power to determine the moral course of one's life and to predict which actions will attract which kind of response, moral appraisal had better look to those aspects of action over which one has most control. The emphasis on the voluntary thus also responds to a concern for freedom as individual self-determination.

This concern is particularly pressing when the code is *legal* rather than moral and the sanctions are state-enforced penal sanctions that constitute a particularly serious threat to individual freedom. By making criminality conditional on voluntariness, sanctions become easier to anticipate from the point of view of the agents. This serves the individual's need for freedom as self-determination: it enhances people's power to determine and predict the course of their lives, particularly with regard to whether they will be subjected to penal sanctions or not. Without those excusing conditions—if people were liable simply for what they did or brought about—they would lose much of their power to determine and predict whether they stayed on the right side of the law. The law's emphasis on voluntariness thus expresses what Hart calls *respect for the individual as a choosing being* (2008b, p. 49). Moreover, it is part of the point of many legal institutions—such as contracts, wills, gifts, and marriages—that they enhance individuals' ability to shape the future: they are legal tools enabling individuals to *lock in* certain outcomes.<sup>23</sup> For the law to interfere in individuals' lives in ways that they could not possibly foresee would therefore defeat one of the very purposes of the law. In the moral and even more so in the legal sphere, the idea of the voluntary thus performs a protective function, helping to respect and defend individual freedom against the claims of society.

<sup>22</sup> This still leaves room for a moral authority that is exerted simply by what one has done—in cases that call for agent-regret as opposed to guilt, for instance; see Williams (1981, 1993, p. 66; 2011, p. 196).

<sup>23</sup> See Hart (2008b, pp. 29–30).

### 3.3 Knowledge and coercion

In a broadly similar spirit, John Hyman has recently argued that “voluntariness is at root an ethical concept” that is “designed for” the purpose of assessing a person’s culpability and “formed by negation,” by “excluding factors that exculpate” (2015, pp. 76–77). Among the factors that are widely taken to exculpate are *ignorance* of what one is doing—sometimes called the “knowledge condition” on responsibility—and doing something under *coercion* (e.g. at the point of a gun)—sometimes called the “freedom condition” on responsibility.<sup>24</sup> When Hyman proposes an *ex negativo* definition of voluntary action as what is “*not* done out of ignorance or compulsion” (2015, p. 77), it is these two conditions he puts front and centre. This raises the question of how the understanding of voluntary agency as acting intentionally and in a normal state of mind relates to the knowledge condition and the freedom condition.

In the case of the knowledge condition, the relation is straightforward—it is already contained in the definition of the voluntary: the distinction between things done in full awareness of their nature and significance and things done out of ignorance is already encapsulated in the requirement that one’s  $\phi$ -ing must be an intentional aspect of what one does, for something can be an *intentional* aspect of what one does only if one is *aware* of it—if one “knows” what one is doing.<sup>25</sup> An action’s being unintentional *is* its being involuntary through ignorance. Of course, as recent debates over moral responsibility have made clear, the relevant notion of knowledge or awareness can be specified in various ways, depending on *how much* it requires one to be aware of and *what kind* of awareness it requires (does one need to *know* or merely *believe* it, and to do so *occurrently* or merely *dispositionally*?).<sup>26</sup> But just as human beings have reasons to discriminate between the intentional and the unintentional that are independent of their concern to allocate responsibility, they are bound to have an interest in *what an agent was aware of* when he or she acted, because that is already crucial to understanding the character of the action, the character of the agent, and the circumstances under which the agent would do something like that again. Did he know what he was doing when he intentionally  $\phi$ -ed? That is, was he aware of the fact that he thereby  $\psi$ -ed? Or—what is sometimes rather more concerning—did he do it even though he did *not* know what it meant and what consequences it would have? Would he have done it anyway if he had known? If we tell him, will he refrain from doing it again? These are questions that human beings have an interest in asking already in virtue of being in the business of interpreting and shaping each other’s actions.

The freedom condition, by contrast, is not yet contained in that definition of the voluntary. Indeed, Williams, like Anscombe, is comfortable describing actions done under coercion as voluntary.<sup>27</sup> He emphasizes that decisions reached under coercion

<sup>24</sup> See, e.g., Fischer and Ravizza (1998, pp. 12–13) and Rudy-Hiller (2018).

<sup>25</sup> This is also how Williams (1995h, pp. 23–26) understand the intentionality requirement.

<sup>26</sup> See Rudy-Hiller (2018) for a helpful overview.

<sup>27</sup> See Anscombe (2008b, p. 127) and Williams (1995h, p. 33n8). In an archival note, however, Anscombe notes that “voluntary behaviour is behaviour in respect of which the behavior is free. To be free

really are *decisions*, characteristically coming out of a process of deliberation issuing in the conclusion that one *has to* do something, because the coercing threat overrides other deliberative priorities.<sup>28</sup> This is not to deny that coercive circumstances can be exculpatory. It is merely to deny that every exculpatory circumstance must register as such by going into forming the voluntary/involuntary distinction. As Williams insists, “the topic of coercion is not part of the theory of action, but of the theory of freedom” (1995h, p. 33n3).

But if a central function of the concept of the voluntary is to separate, in the name of fairness and freedom, actions that are *more* from actions that are *less* under the agent’s control, then perhaps we do, after all, have reason to refine the generic notion of the voluntary we started out from to exclude coerced action. There is an undeniably important sense in which the agent who acts under threat of serious harm is deprived of control over the action, and the intentions expressed by the action are not really *the agent’s* at all, but the coercer’s. Like acting in an abnormal state of mind, this dissociates the action from the agent. And indeed, legal codes tend to treat the fact that an action was done under *duress* (coercion by the threat of serious harm) as an excusing condition; and Hart, for example, characterizes coerced actions as *involuntary* (2008c, pp. 143–44; 2008d, p. 14). To better serve the concerns that fuel our interest in it, therefore, the notion of the voluntary should be understood more restrictively:

A  $\phi$ -s *voluntarily* if and to the extent that  $\phi$ -ing is an intentional aspect of an action that A performs uncoerced and in a normal state of mind.

One might protest on Williams’s behalf that this introduces a moral element from the theory of freedom into what was supposed to be, in the first instance, a psychological idea in the theory of action, on a par with ideas of choice, decision, belief, or desire. But even the earlier, more permissive definition of the voluntary could hardly escape drawing on a morally laden understanding of what kinds of states of mind appropriately attract appraisals of responsibility. Moreover, if the reasons we have the idea of the voluntary in the first place, over and beyond the ideas of acting intentionally and in a normal state of mind, are *moral* reasons, then it is only right that these moral concerns should also substantively *shape* the resulting idea of the voluntary. In discussions of thick concepts, the point is often made that a thick concept’s extension is a function of our evaluative interests in deploying the concept, and not specifiable independently of them. One might argue that the same is true of the concept of the voluntary. When taken in isolation from the concept of the voluntary and the moral concerns that motivate its use, its constituent notions, such as that of a normal state of mind, are unlikely to have the same extensions as when deployed alongside other constituent notions in the service of moral concerns. This is why the concerns for fairness and freedom call for a genuinely new concept

---

Footnote 27 continued

is to be in a situation of possibility of determining something to accord with one’s will” (Archive, Box 9, File 304, p. 1; cited in Bierson and Schwenkler (forthcoming, §3)).

<sup>28</sup> See Williams (1995c, p. 5; 1995h, pp. 33n3, 33n8).

that is more than the sum of its constituent concepts. The notions that coalesce in the concept of the voluntary need to be understood in a certain way and be appropriately linked, and neither these constituent notions nor their appropriate link are specifiable independently of the moral interests that motivate, guide, and inform the use of the concept of the voluntary.

At the same time, Williams and Anscombe are right to emphasize that the idea of the voluntary is, in the first instance, a concept in the theory of action. It is a psychological idea—only one that acts as a linchpin between psychological and moral ideas, connecting ideas of action, deliberation, intention, decision, and will with ideas of praise, blame, responsibility, and exculpation. What the idea of the voluntary *picks out* is a collection of properties of actions, which properties they have in virtue of their relation to the agent's deliberation and state of mind; but the *reasons* for picking out just these properties and grouping them as constituents of a significant property—voluntariness—are *moral* reasons, and the consequences of displaying or lacking that property are *moral* consequences. Hence, the idea of the voluntary marks a place where ideas about mind and action make contact with moral pressures. In a hybridizing classification that will prove helpful below in understanding the “moralization of psychology,” we might say that the idea of the voluntary is a psychological idea that serves and is shaped by moral concerns.

Finally, besides explaining and shaping the concept of the voluntary, the concerns for fairness and freedom also *vindicate* our use of it. They yield a basis on which to argue that the concept of the voluntary does important work by serving these concerns. And, crucially, this is work the concept already does in its superficial form, so that there is nothing here to motivate deepening it.

#### 4 Why the voluntary is deepened

The upshot so far has been that we need the concept of the voluntary in *some* form because we feel the need, for the sake of fairness and freedom, to exempt individuals from some of their moral and legal responsibility for some of their involuntary acts. The question we now turn to is where the pressure to deepen the idea of the voluntary comes from.

The answer I want to explore is that the pressure to deepen the idea of the voluntary stems from the moral demand to make attributions of responsibility not just *fairer* than they would be without the idea of the voluntary, but *ultimately* fair. The superficial idea of the voluntary goes some way towards meeting the demand for fairness, but by no means all the way. Letting moral responsibility rest largely on a superficial notion of the voluntary helps exempt the agent from responsibility for things done through sheer bad luck. But it still leaves a great deal of residual unfairness, because *whether one is in a position* voluntarily to do the right thing itself remains a matter of luck—a matter of one's inherited disposition, upbringing, socialization, or incidental confrontation with certain situations. Appraisals of responsibility based on a superficial idea of voluntariness are thus fairer than they would be if they completely ignored the voluntary/involuntary distinction, but they are not *ultimately* fair. The superficial idea of the voluntary falls foul of the demand

for ultimate fairness that Michael Zimmerman forcefully expresses when he insists that “the degree to which we are morally responsible cannot be affected by what is not in our control. Put more pithily: luck is irrelevant to moral responsibility” (2002, p. 559).

The moral demand for responsibility to be attributed on an ultimately fair basis means that the idea of the voluntary comes under pressure to be deepened in such a way as to accommodate this demand for ultimate fairness. For, given a deepened idea of the voluntary, that which is a product and expression of the agent’s will can be fully isolated from that which is a product of forces beyond the agent’s control.

The result is a problematic moralization of psychology: it is one thing for a psychological idea such as the idea of the voluntary to tie in with and be refined by moral demands—this in itself may be an unproblematic moralization of psychology; but the moralization of psychology becomes *problematic* when it goes further than that, and the psychological ideas are warped under moral pressure to the point where they become *psychologically unrealistic*, which is to say when they become inconsistent with the rest of what we take ourselves to know about human psychology.

This is again a point at which we can usefully turn to Williams, because the idea that psychology can become problematically moralized under the pressure of moral demands is a thread that runs through much of his oeuvre. Already in 1963, Williams makes the point—which he credits to Iris Murdoch—that it is an “evaluatively motivated picture of the mind” which “sharply distinguishes between ‘reason’ and ‘will’” (1963, p. 136). In *Shame and Necessity*, he then suggests that it was Plato who first “ethicized psychology” with his tripartite model of the soul, because he defined “the functions of the mind, especially with regard to action ... at the most basic level in terms of categories that get their significance from ethics” (1993, p. 160). In particular, Plato’s stark division between “rational concerns that aim at the good, and mere desire” (1993, p. 42) allowed him to introduce a “featureless moral self” (1993, p. 160) into his psychology, a locus of agency that remained uncontaminated by contingent desires. What makes this an example of a problematically moralized psychology is not that it plays a morally significant role, or even that it draws on values—a realistic psychology, as Williams notes, need not be “value-free”; but it “leaves it open, or even problematical, in what way moral reasons and ethical values fit with other motives and desires, how far they express those other motives, and how far they are in conflict with them” (1995g, p. 202).<sup>29</sup> What a problematically moralized psychology lacks is precisely this openness: it closes off the very possibility of ineliminable conflict between the psychological and the moral. Psychological ideas are distorted to ensure that moral demands can be met.

The deepening of the psychological idea of the voluntary, I want to suggest, is a central example of such a problematic moralization of psychology. To satisfy the moral demand for ultimate fairness, voluntariness cannot reside at the superficial level at which any expression of the will has already been refracted through

<sup>29</sup> See also Williams (2006c, p. 78).

contingent circumstances. To form an *ultimately fair* basis of ascriptions of responsibility, the idea of the voluntary must meet two demands:

*Equal Opportunity*: it has to be the idea of something that agents have an *equal opportunity* to exercise, for instance by being grounded in a capacity that any agent possesses already merely in virtue of being a rational agent;

*Total Control*: it has to be the idea of something that any agent with the relevant capacity has *total control* over, which means that it has to be extricated from any contingent circumstances that might constrain or predetermine its course.<sup>30</sup>

It is not hard to see how these two demands for something that is perfectly evenly distributed and entirely within the agent's control would drive the deepening of the idea of the voluntary. As long as the idea of the voluntary falls short of these two conditions, it leaves the agent at the mercy of some measure of contingency and therefore of unfairness. To meet those conditions, voluntariness has to be located beyond any determination by empirical contingencies, at some deeper level which the theory doing the deepening obligingly provides.

One possible explanation for why many theories of action have attempted to deepen the idea of the voluntary, then, is that these theories are morally motivated by a yearning for life to be ultimately fair. Assessing to what extent this hypothesis can be corroborated once we examine the thoughts and motivations of philosophers such as Descartes, Locke, Hume, or Mill lies beyond the scope of this paper. My concern here is with the *systematic connection* between the deepened idea of the voluntary and the yearning for ultimate fairness, not with the historical question of when and by whom that connection has in fact (self-consciously or not) been exploited in the course of history. The systematic point is that by deepening the idea of the voluntary in such a way that voluntary agency lies beyond contingent empirical determination, an account of the psychology of agency acquires the right shape to hold out the promise of allocating responsibility and blame on an ultimately fair basis. Deepening the idea of the voluntary promises to turn it into a shelter from contingency.

## 5 The need to resist the pressure to deepen: a three-tier model

We come now to the negative argument as to why the idea of the voluntary should not be deepened. In Sect. 3, we saw that the reason we need *some* such concept in the first place is that it works in the service of fairness and freedom by exempting individuals from some of their moral and legal responsibility for some of their

---

<sup>30</sup> Again, the notion of *total control* can be spelled out in different ways depending on the theory of action at issue. If total control is theorized in terms of having a real choice between hitherto objectively undetermined courses of action, for example, my really choosing to  $\phi$  at time  $t$  must itself be unconditioned by prior circumstances *and* objectively determine the  $\phi$ -ing to happen, where, up to  $t$ , it was still possible for the  $\phi$ -ing not to happen. See Sect. 2 and note 13 for alternative but functionally equivalent ways of theorizing the notion of total control.

involuntary acts. As we saw in Sect. 4, however, this still leaves us with considerable residual unfairness, since whether one is in a position voluntarily to do the right thing remains a matter of luck. The concern to eliminate this residual unfairness might then be thought to motivate the deepening of the idea of the voluntary. But, as I shall argue in this section, far from doing the same job better, a deepened concept of the voluntary ceases to do the work it did in its superficial form and lands us in broader trouble.

The source of that trouble is that if some psychological idea is problematically moralized under the pressure of moral demands—in particular, if the idea of the voluntary is deepened under the pressure of a demand for ultimate fairness—this generates a pressure on the rest of our psychological ideas to support that moralized idea and a pressure on our physical or metaphysical ideas to support such a psychology. The relevant dynamics can be described in terms of a simple three-tier model. It organizes our moral, psychological, and physical ideas (in the broad sense of “physical” that includes biological, physiological, and metaphysical ideas, as it did in the ancient notion of *physis*) into three vertically layered tiers of ideas, with the physical tier at the bottom, the psychological tier in the middle, and the moral tier at the top. Each tier comprises ideas such as the following:

*Moral Ideas*: moral responsibility, blame, praise, justice, fairness, freedom, etc.

*Psychological Ideas*: voluntariness, will, trying, choice, decision, action, belief, desire, etc.

*Physical Ideas*: bodily movements, cause, effect, determinism, quantum randomness, etc.

This three-tier model takes us beyond the traditional two-tiered framing of the free will problem as a matter of whether our *moral* ideas are compatible with *physical* ideas such as determinism or randomness at the quantum level. The three-tier model makes the framing of the problem in terms of the possibility of reconciling moral responsibility with determinism look like a “structural misconception” (1995c, p. 6), as Williams puts it; the three-tier model brings out that, *in the first instance*, “our ideas of blame and responsibility are answerable to an adequate psychology (rather than to generic worries about determinism)” (1995d, p. 45n10). Accordingly, there are not one, but *two* points at which the question of the relation of one set of concepts to another set arises, and as a result, we get not one, but *two* free will problems: (1) How can our physical ideas be reconciled with our psychological ideas? And (2) how can our psychological ideas be reconciled with our moral ideas?

What this simplifying model illustrates is that securing a fit between moral ideas and psychological ideas by deepening the idea of the voluntary comes at the price of exerting pressure on psychological ideas in a way that may bring them into tension with physical ideas: if the will expressed in voluntary action is to be a force that is entirely pure of any conditioning influence by unfairly distributed contingencies, how are we to make sense of such a force in physical terms?

To resolve the tensions that problematically moralized psychological ideas generate in relation to physical ideas, the moralization has to *seep through* or *extend to* the physical. That is to say, one's conception of nature or its metaphysical underpinnings has to be adapted to meet the moral demand for some deepened idea of voluntary agency that can provide an ultimately fair basis for responsibility. Aristotle's biology, with its idea that the proper natural development of the human animal issues in virtue, can be thought of as exemplifying a conception of nature that is tailored to ensure that moral demands can be met. Similarly, Plato's dualistic metaphysics of the soul and his theory of the Forms support his tripartite psychology in just the way required to secure the realizability of his moral ideals. But perhaps the best example of a moralized metaphysics tailored to support moral demands that are specially focused on the idea of the voluntary is Kant's postulation of the noumenal realm as the locus of the transcendental subject's unconditioned will.<sup>31</sup>

On a picture on which a moralized psychology finds support in physical ideas, there can be a harmonious fit between the fundamental nature of reality, the operations of human psychology, and the demands of morality. There is considerable attraction in such a picture. It holds out the promise that everyone will be offered an equal and fair opportunity to live up to the demands of morality: if the moral shape of one's life depends exclusively on what one voluntarily does; if the voluntary/involuntary distinction is applied at the deeper level of the purely rational self rather than at the superficial level of the contingently constituted empirical self; and if one's metaphysical view of things supports the existence of such a deeper level; then the moral shape of one's life really can be entirely within one's control and reassuringly sheltered from luck.

But, on a thoroughly disenchanted picture of the world, it is far from clear that a moralized psychology can find the support it requires in our physical ideas. The story of the rise of modern science is also the story of how our physical ideas have become *demoralized*, in the sense of becoming more emancipated from moral demands and less accommodating of them. As a result, the deepened idea of the voluntary at work in the theory of volitions or in the Kantian theory of the unconditioned will of the rational agent no longer neatly dovetails with our physical and metaphysical ideas, and it comes to look as though the idea of a voluntary act is never fully instantiated. Absent some layer of reality beyond the empirically conditioned phenomenal or material world, our modern physical ideas do not obviously leave room for pure volition or an unconditioned will, and without some such force, the idea of the voluntary cannot realistically be deepened; there cannot then be an action that is voluntary all the way down, unconditioned by any contingent factor lying beyond the control of the agent's will. Hence, the demoralization of physical ideas creates a tension with moralized psychological ideas and renders them problematic, in particular by making it hard to see how human agency can live up to that conception of it.

<sup>31</sup> Though some have insisted that even Kant ultimately does not deny that moral action is subject to contingent empirical determination; see Heyd (1997) and Hartman (2019).

The effect of combining in people's minds a deepened idea of the voluntary with physical ideas that deny its instantiation is not *necessarily* to produce the conviction that no act is ever voluntary. People are quite capable of holding on, at once, to two sets of ideas that are in tension with each other without realizing it. But it does *set them up* to conclude, when they reflect on the voluntariness of a particular action and inquire more deeply into its aetiology, that this action was not *really* voluntary, because it reflected the agent's genetic predispositions, or upbringing, or some other contingent circumstantial factor. On this basis, they may come to realize that something similar holds true of every action they can point to. *Pace* the compatibilists, they would then not be mistaken about the incompatibility of their physical ideas with their idea of voluntariness, but would correctly have grasped the implications that a deepened idea of the voluntary has when set against a naturalized or disenchanting conception of the world. If one's idea of a voluntary action requires the action to be pure of contingency or luck in every respect, sufficiently close scrutiny will eventually disqualify any action. When set against a disenchanting conception of the world, a deepened idea of the voluntary must see its extension shrink to the point of becoming empty. And insofar as its extension is empty, the idea of the voluntary loses its applicability to our actual experience and fails to mark out any actions as appropriate objects for our practices of responsibility attribution. Perhaps some concepts are worth having even if they are never satisfied, but insofar as the idea of the voluntary earns its keep by rendering us sensitive to *differences* between human actions of the kind actually to be met with, it is not one of those concepts. Its failure to carve off any actual action as voluntary renders the idea pointless—a distinction without a difference.

At the outset, we saw Williams remark that while the notion of the voluntary may be "extended or contracted in various ways," it can hardly be deepened, because the effect of trying to do so is "to put it beyond all recognition." We can now see exactly what this formulation might be getting at. Users of a concept of the voluntary can extend or contract the concept by extending or contracting the set of actions they are prepared to count as voluntary, thereby adjusting the concept to their purposes, to certain domains of application, or to the peculiarities of their cultural situation. We contracted the concept in this very discussion, for example, by adjusting it to exclude coerced action. Such adjustments to the concept's scope are unproblematic—as, indeed, are adjustments at the object level, whereby the range of what is voluntary is itself reduced or expanded without changing the concept of the voluntary, for instance by learning to bring movements under the control of one's will that one previously lacked control over (as infants learning to master their bodies do).

But trying to deepen the concept of the voluntary has the effect of putting it beyond all recognition in that it makes it impossible ever to *recognize*, in the disenchanting, contingent muck of human affairs, the metaphysically deep, "utter voluntariness" that the theory doing the deepening encourages one to look for. And once disappointed, these expectations invite the conclusion that "without ...

utter voluntariness, there is only force” (Williams, 2011, p. 218).<sup>32</sup> A form of scepticism about free will then beckons, as it looks like no-one ever really acts voluntarily.

What makes this kind of scepticism specially interesting is its familiarity. It derives not from an arcane theory concocted in a seminar room and easily forgotten outside it, but from a moral demand the tug of which just about anyone can recognize. We are familiar with situations in which deeper scrutiny of the conditions that led up to a superficially voluntary action has people revise their perception of the action as voluntary and withhold blame as a result. *Tout comprendre, c'est tout pardonner*, as the old adage has it—to understand all is to excuse all. Insofar as one comes to view an action, even if only temporarily, in terms of a deepened concept of the voluntary, scrutiny of the action's origins in the absence of Platonic or Kantian metaphysical theories lending succour to a deepened concept of the voluntary will lead the concept of the voluntary to lose its grip on the action, because people's control over who they become, what their characters are, what they live through, and what they end up having to do, will not seem to go far enough to warrant the application of the concept. This, presumably, is what Williams means when he gnomically suggests that if we push beyond a certain point the question whether an act was voluntary, “we sink into the sands of an everyday, entirely justified, skepticism,” one that “is indeed everyday, generated by an honest acquaintance with human affairs” (1993, p. 67).

But if this scepticism really is everyday, then its practical relevance will be magnified by the fact that the idea of the voluntary is a central load-bearing notion for our moral and legal outlook. The idea of the voluntary, as we saw, is a linchpin between the psychology of action and the normative demands we make on actions in the moral and legal sphere. If we assume, as has often been argued,<sup>33</sup> that these normative demands by which a society regulates the behaviour of its members are indispensable to society's functioning, deepening the idea of voluntariness threatens to rob these socially indispensable practices of responsibility attribution of their grip on human affairs, and hence also of their efficacy in helping us live together.

Crucially, however, it is only insofar as the idea of the voluntary is deepened that the “demoralization” of physical ideas generates what we identified above as the *first* problem of free will, for only then will the idea of the voluntary come into conflict with physical ideas such as determinism. In its superficial form, the idea generates no such conflict, as suggested for instance by Brian O'Shaughnessy's reconciliation of causal determinism with voluntary action, which Williams explicitly endorses.<sup>34</sup> That is why Williams writes that “there is a problem of free will only for those who think that the notion of the voluntary can be

<sup>32</sup> It is at this point that the critique of deepened notions of voluntariness connects with Williams's critique of the “morality system.” See Loudon (2007, pp. 110–11) and Queloz (forthcoming). See also Hartman (2016) for a congenial argument directed against luck-free conceptions of moral responsibility.

<sup>33</sup> See, e.g., Strawson (2008), Russell (manuscript), Fricker (2016, 2019, forthcoming), and McGeer (2019).

<sup>34</sup> See O'Shaughnessy (2008a, ch. 4; 2008b, esp. chs. 11 and 17). For Williams's endorsements, see Williams (1995a, p. 579; 1995c, p. 8).

metaphysically deepened” (1993, p. 68), and part of what drives Williams’s turn to the Pre-Socratic Greeks is the desire to recover—or remind us of the point and value of—an idea of the voluntary as yet unmarked by the moral pressure to deepen it.

## 6 Balancing moral and practical demands

The reason the idea of the voluntary should not be deepened is therefore that it is *essentially* superficial: it is *only* if the idea is understood superficially that it can do its work for us by rendering us sensitive to differences *within* our experience. Only if we resist the temptation to deepen it can it get a grip on the disenchanted world we live in, because the effect of deepening it is to contract its extension to the point where no action qualifies as truly voluntary. The idea of the voluntary cannot be deepened without losing its applicability to the kind of world we take ourselves to live in.

It follows that if the idea of the voluntary (along with the appraisals of responsibility that build on it) is to do any practical work for us in a world in which every action displays an element of contingency, its extension within that world had better not shrink to the point of becoming empty: the contrast between the voluntary and the involuntary had better remain an internal contrast *within* our experience. If our moral and legal ideas are to be applicable to us contingently determined agents, the idea of the voluntary that focuses their application must accommodate some measure of contingency. (This is not to lower our sights because we cannot, as agents, completely extricate ourselves from the influence of contingent forces; far from always being a limitation to be overcome, after all, contingent influences are also what enables an agent to develop an individual character and a distinctive identity in the first place.)

To retain the superficiality of the voluntary/involuntary distinction is precisely to retain this ability to draw a contrast within the world we live in, by contrasting, not conditioned with unconditioned actions, but different kinds of conditioned actions. Whereas deepened ideas of the voluntary conceive of voluntariness and contingency as mutually exclusive, superficial ideas are more tolerant of contingency, drawing the voluntary/involuntary distinction *within* a range of actions that are *all* adulterated with contingency to some degree, but adverting instead, and rather more helpfully, to the differences in *how much* contingency is involved in them. A superficial voluntary/involuntary distinction contrasts not the pure white of voluntariness with the pitch black of contingency, but the grey nuances that result from their being mixed in varying proportions. It thus renders us, and our practices of responsibility attribution, sensitive to the *degree* to which a conditioned action can be attributed to the agent’s will.

This is something that a deepened notion of voluntariness as defined in Sect. 2 cannot do for us, because it is not gradable—it is a dichotomous rather than continuous concept, which categorically separates voluntary action from every other kind of action or happening, and, by holding up unconditioned purity from contingency as an ideal, levels the differences between conditioned actions.

A superficial notion of voluntariness, by contrast, *can* be—though, as the criminal law cases forcing a binary distinction remind us, it need not always be—a gradable notion: the less an action is the product of forces external to the agent’s will, the more voluntary it is, and the more the agent deserves to be held accountable for it. Such an idea of the voluntary enables us to accept that even the most fully voluntary action reflects influences beyond the control of the agent’s will while still retaining its differential applicability to our actions; it thereby serves the need to hold people accountable for at least some of their actions while also serving the concerns for freedom and fairness as far as realistically possible.

The kind of concept of the voluntary we need, then, is one that balances the moral demands for freedom and fairness against the practical demands to hold people to account and to be realistic about the pervasiveness of contingency: it should render us sensitive to differences in the degree to which actions are under the control of the agent’s will while retaining the wide applicability of our regulatory practices of responsibility attribution by accommodating the fact that even the most voluntary actions are still conditioned by contingent factors lying beyond the agent’s control.

## 7 Conclusion

Contrary to what it may sound like at first, the thesis of the essential superficiality of the idea of the voluntary does not highlight a flaw, but a feature worth preserving; and far from being a metaphysical thesis about the essence of voluntary, it is a thesis in “conceptual ethics”—a contribution to critical reflection on what kind of concept of the voluntary we have most reason to use.<sup>35</sup> I have argued that the forms in which the concept of voluntariness is worth having are those in which we refrain from turning it into a metaphysically deep idea, because for us, any attempt to deepen the idea must issue in scepticism about the possibility of voluntary action. As long as it remains superficial, the idea of the voluntary does little to answer traditional worries about free will; but it also avoids raising those worries in the first place, and thereby avoids threatening the legitimacy of our appraisals of responsibility across the board; and as long as it remains superficial, the idea of the voluntary does important work for us by enabling us to heed claims of fairness and individual freedom in our appraisals of responsibility as far as the pervasiveness of contingency will allow.

Where does this leave us? In the end, it still leaves us facing the *second* problem of free will, the problem of how to reconcile our psychological with our moral ideas. It is in this second reconciling project that Williams sees a major remaining difficulty for compatibilism, and this is why he is not a complacent compatibilist:

Can the reconciling project succeed? Between determinism (or as much naturalistic explanation as you like), and relevant psychological concepts, yes. Between both of these, and the ethical conceptual scheme, no, not as it stands. (Williams, 1995c, p. 19)

---

<sup>35</sup> See Burgess and Plunkett (2013a, 2013b).

Even if we thoroughly demoralize our psychological ideas and succeed in reconciling them with our physical ideas, there remains a tension between our psychological ideas about human action and our moral ideas as long as the latter include a demand for *ultimate* fairness. In particular, there remains a tension between the superficial idea of the voluntary and the moral demand that the actions we are responsible for—the actions which determine the moral shape of a life—must be entirely free of luck.

If what I have argued is along the right lines, however, the way to resolve this residual tension between our psychological and moral ideas is not to deepen our idea of the voluntary so as to shelter voluntary agency from luck. We need our psychological *and* our moral ideas to accommodate the fact that our lives and our actions are infused with contingency. Cultivating and defending a superficial idea of the voluntary is an important step in this direction, since the idea's superficiality allows it to accommodate contingency while still doing its moral work, channelling attributions of responsibility away from those actions that are least within the agent's control. But it can only do this work if it resists the moral pressure for ultimate fairness and remains superficial. The idea of the voluntary thus turns out to be superficial for good reason—as the Nietzschean *bon mot* has it, it is superficial out of profundity.<sup>36</sup>

**Acknowledgements** I am indebted to Adrian Moore, Agata Łukomska, András Szígeti, Anton Emilsson, Benjamin De Mesel, Brian Leiter, Caroline Touborg, Constantine Sandis, Damian Cueni, Geraldine Ng, Gideon Rosen, Gunnar Björnsson, John Hyman, Jonathan Dancy, Lorenzo Greco, Marcel van Ackeren, Markus Wild, Matthew Talbert, Miranda Fricker, Patricia Williams, Paul Russell, Philipp Schwind, Robert Hartman, Sophie-Grace Chappell, and Stephen Darwall for valuable discussions of these issues. I have also benefited enormously from the expertise and the generous comments of two anonymous reviewers. My work on this paper was supported by Grant No. P400PG\_199210 of the Swiss National Science Foundation and by Wolfson College, Oxford.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

Alanen, L. (2002). Descartes on the will and the power to do otherwise. In H. Lagerlund & M. Yrjönsuuri (Eds.), *Emotions and choice from Boethius to Descartes* (pp. 279–298). Springer.

---

<sup>36</sup> See Nietzsche (2001, Preface, §4; 2005, Epilogue, §2), who uses the phrase in a different connection. Williams (1993, p. 68), however, applies the remark to—among other things—the voluntary, which further supports my reading on which he saw the superficiality of the idea not as a defect, but as a strength.

- Alvarez, M., & Hyman, J. (2019). Philosophy of action 1945–2015. In K. Becker & I. D. Thomson (Eds.), *The Cambridge history of philosophy, 1945–2015* (pp. 103–114). Cambridge University Press.
- Anscombe, G. E. M. (1957). *Intention*. Blackwell.
- Anscombe, G. E. M. (2005). Action, intention and ‘double effect.’ In M. Geach & L. Gormally (Eds.), *Faith in a hard ground: Essays on religion, philosophy and ethics by G.E.M. Anscombe* (pp. 207–226). Imprint Academic.
- Anscombe, G. E. M. (2008a). On being in good faith. In M. Geach & L. Gormally (Eds.), *Faith in a hard ground: Essays on religion, philosophy and ethics by G.E.M. Anscombe* (pp. 101–112). Imprint Academic.
- Anscombe, G. E. M. (2008b). Sin: the McGivney lectures. In M. Geach & L. Gormally (Eds.), *Faith in a hard ground: Essays on religion, philosophy and ethics by GEM Anscombe* (pp. 117–156). Imprint Academic.
- Bierson, M., & Schwenkler, J. (forthcoming). What is the bearing of thinking on doing? In A. Haddock & R. Wiseman (Eds.), *The Anscombean mind*. Routledge.
- Burgess, A., & Plunkett, D. (2013a). Conceptual ethics I. *Philosophy Compass*, 8(12), 1091–1101.
- Burgess, A., & Plunkett, D. (2013b). Conceptual ethics II. *Philosophy Compass*, 8(12), 1102–1110.
- Candlish, S., & Damjanovic, N. (2013). Reasons, actions, and the will: The fall and rise of causalism. In M. Beaney (Ed.), *The Oxford handbook of the history of analytic philosophy* (pp. 689–708). Oxford University Press.
- Crisp, R. (2017). Moral luck and equality of opportunity. *Aristotelian Society Supplementary*, 91(1), 1–20.
- D’Oro, G., & Sandis, C. (Eds.). (2013). *Reasons and causes: Causalism and anti-causalism in the philosophy of action*. Palgrave Macmillan.
- Davidson, D. (1980). *Essays on actions and events*. Clarendon Press.
- De Mesel, B. (manuscript). Being and holding responsible: reconciling the disputants through a meaning-based strawsonian account.
- Deigh, J. (2008). *Emotions, values, and the law*. Oxford University Press.
- Descartes, R. (1964–1976). *Oeuvres de Descartes* (Vol. I–XI). Vrin.
- Duff, A. (2004). Action, the act requirement and criminal liability. In J. Hyman & H. Steward (Eds.), *Agency and action. Royal Institute of Philosophy Supplements* (Vol. 55, pp. 69–103). Cambridge University Press.
- Duff, A., & von Hirsch, A. (1997). Responsibility, retribution and the voluntary: A response to Williams. *Cambridge Law Journal*, 56(1), 103–113.
- Emilsson, A. (manuscript). The unanswered question of ‘freedom and resentment’.
- Fischer, J. M. (1994). *The metaphysics of free will: An essay on control*. Blackwell.
- Fischer, J. M., & Ravizza, M. (1998). *Responsibility and control: A theory of moral responsibility*. Cambridge University Press.
- Fricker, M. (forthcoming). *Blaming and forgiving: The work of morality*. Oxford University Press.
- Fricker, M. (2016). What’s the point of blame? A paradigm based explanation. *Nouûs*, 50(1), 165–183.
- Fricker, M. (2019). Forgiveness: An ordered pluralism. *Australasian Philosophical Review*, 3(1), 241–260.
- Glock, H.-J. (2014). Reasons for action: Wittgensteinian and Davidsonian perspectives in historical, meta-philosophical and philosophical context. *Nordic Wittgenstein Review*, 3(1), 7–46.
- Hart, H. L. A. (1963). Acts of will and legal responsibility. In D. Pears (Ed.), *Freedom and the will* (pp. 38–47). Palgrave.
- Hart, H. L. A. (2008a). Acts of will and responsibility. In H. Hart (Ed.), *Punishment and responsibility: Essays in the philosophy of law* (2nd ed., pp. 90–112). Oxford University Press.
- Hart, H. L. A. (2008b). Legal responsibility and excuses. In *Punishment and responsibility: Essays in the philosophy of law* (2nd ed., pp. 28–53). Oxford University Press.
- Hart, H. L. A. (2008c). Negligence, mens rea, and criminal responsibility. In *Punishment and responsibility: Essays in the philosophy of law* (2nd ed., pp. 136–157). Oxford University Press.
- Hart, H. L. A. (2008d). Prolegomenon to the principles of punishment. In *Punishment and responsibility: Essays in the philosophy of law* (2nd ed., pp. 1–27). Oxford University Press.
- Hartman, R. J. (2016). Against luck-free moral responsibility. *Philosophical Studies*, 173(10), 2845–2865.
- Hartman, R. J. (2019). Kant does not deny resultant moral luck. *Midwest Studies in Philosophy*, 43(1), 136–150.

- Heyd, D. (1997). Moral and legal luck: Kant's reconciliation with practical contingency. *Annual Review of Law and Ethics*, 5, 27–42.
- Hornsby, J. (1980). *Actions*. Routledge.
- Hyman, J. (2011). Action and the will. In O. Kuusela & M. McGinn (Eds.), *The Oxford handbook of Wittgenstein* (pp. 451–471). Oxford University Press.
- Hyman, J. (2015). *Action, knowledge, and will*. Oxford University Press.
- James, W. (1981). *The principles of psychology* (Vol. II). Harvard University Press.
- Jayasekera, M. (2016). Responsibility in Descartes's theory of judgment. *Ergo*, 3(12), 321–347.
- Kenny, A. J. P. (1972). Descartes on the will. In R. J. Buter (Ed.), *Cartesian studies* (pp. 1–31). Blackwell.
- Krishnan, N., & Queloz, M. (manuscript). The shaken realist: Bernard Williams, the war, and philosophy as cultural critique.
- La Boétie, E. (2016). *Discours de la servitude volontaire*. Flammarion.
- Levy, N. (2011). *Hard luck: How luck undermines free will and moral responsibility*. Oxford University Press.
- Louden, R. B. (2007). The critique of the morality system. In A. Thomas (Ed.), *Bernard Williams* (pp. 104–134). Cambridge University Press.
- McGeer, V. (2019). Scaffolding agency: A proleptic account of the reactive attitudes. *European Journal of Philosophy*, 27(2), 301–323.
- McKenna, M., & Coates, D. J. (2021). Compatibilism. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (2021st ed.). Stanford University.
- Mele, A. R. (2017). *Aspects of agency: Decisions, abilities, explanations, and free will*. Oxford University Press.
- Moore, A. W. (2003). Williams on ethics, knowledge and reflection. *Philosophy*, 78, 337–354.
- Moore, A. W. (2006). Bernard Williams: Ethics and the limits of philosophy. In J. Shand (Ed.), *Central works of philosophy volume 5: The twentieth century: quine and after* (Vol. 5, pp. 207–226). Acumen.
- Moore, M. S. (2010). *Act and crime: The philosophy of action and its implications for criminal law*. Oxford University Press.
- Nietzsche, F. (2001). *The gay science: With a Prelude in German rhymes and an appendix of songs* (J. Nauckhoff, & A. Del Caro, Trans.). Cambridge University Press.
- Nietzsche, F. (2005). Nietzsche Contra Wagner. In A. Ridley & J. Norman (Eds.), *The anti-christ, ecce homo, twilight of the idols, and other writings* (pp. 263–282). Cambridge University Press.
- O'Shaughnessy, B. (1973). Trying (as the mental 'pineal gland'). *The Journal of Philosophy*, 70(13), 365–386.
- O'Shaughnessy, B. (2008a). *The will: A dual aspect theory* (Vol. I). Cambridge University Press.
- O'Shaughnessy, B. (2008b). *The will: A dual aspect theory* (Vol. II). Cambridge University Press.
- Pereboom, D. (2001). *Living without free will*. Cambridge University Press.
- Queloz, M. (forthcoming). A shelter from luck: The morality system reconstructed. In A. Szigeti & M. Talbert (Eds.), *Morality and agency: Themes from Bernard Williams*. Oxford University Press.
- Queloz, M. (2018). Williams's pragmatic genealogy and self-effacing functionality. *Philosophers' Imprint*, 18(17), 1–20.
- Queloz, M. (2021a). *The practical origins of ideas: Genealogy as conceptual reverse-engineering*. Oxford University Press.
- Queloz, M. (2021b). The self-effacing functionality of blame. *Philosophical Studies*, 178(4), 1361–1379.
- Rosen, M. (2013). *On voluntary servitude: False consciousness and the theory of ideology*. Wiley.
- Rudy-Hiller, F. (2018). The epistemic condition for moral responsibility. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (2018 ed.). Stanford University.
- Russell, B. (1921). *The analysis of mind*. George Allen and Unwin.
- Russell, P. (manuscript). The limits of responsibility skepticism.
- Russell, P. (forthcoming). Free will and the tragic predicament: Making sense of Williams. In A. Szigeti & M. Talbert (Eds.), *Morality and agency: Themes from Bernard Williams*. Oxford University Press.
- Russell, P. (2017a). Free will pessimism. In D. Shoemaker (Ed.), *Oxford studies in agency and responsibility* (Vol. 4, pp. 93–120). Oxford University Press.
- Russell, P. (2017b). *The limits of free will: Selected essays*. Oxford University Press.
- Ryle, G. (2009). *The concept of mind*. Routledge.
- Saunders, K. W. (1988). Voluntary acts and the criminal law: Justifying culpability based on the existence of volition. *University of Pittsburgh Law Review*, 49(2), 443–476.

- Schneck, A. C. (2019). Elisabeth of Bohemia's neo-peripatetic account of the emotions. *British Journal for the History of Philosophy*, 27(4), 753–770.
- Strawson, P. F. (2008). Freedom and resentment. *Freedom and resentment and other essays* (pp. 1–28). Routledge.
- Waller, B. N. (2011). *Against moral responsibility*. MIT Press.
- Williams, B. (1963). Postscript. In D. Pears (Ed.), *Freedom and the will* (pp. 105–137). Palgrave.
- Williams, B. (1981). Moral luck. *Moral luck* (pp. 20–39). Cambridge University Press.
- Williams, B. (1993). *Shame and necessity*. University of California Press.
- Williams, B. (1995a). Ethics. In A. C. Grayling (Ed.), *Philosophy 1: A guide through the subject* (pp. 545–582). Oxford University Press.
- Williams, B. (1995b). Formal and substantial individualism. *Making sense of humanity and other philosophical papers, 1982–1993* (pp. 123–134). Cambridge University Press.
- Williams, B. (1995c). How free does the will need to be? *Making sense of humanity and other philosophical papers, 1982–1993* (pp. 3–21). Cambridge University Press.
- Williams, B. (1995d). Internal reasons and the obscurity of blame. *Making sense of humanity and other philosophical papers, 1982–1993* (pp. 35–45). Cambridge University Press.
- Williams, B. (1995e). Moral luck: A postscript. *Making sense of humanity and other philosophical papers, 1982–1993* (pp. 241–247). Cambridge University Press.
- Williams, B. (1995f). Nietzsche's minimalist moral psychology. *Making sense of humanity and other philosophical papers, 1982–1993* (pp. 65–78). Cambridge University Press.
- Williams, B. (1995g). Replies. In J. E. J. Altham & R. Harrison (Eds.), *World, mind, and ethics: Essays on the ethical philosophy of Bernard Williams* (pp. 185–224). Cambridge University Press.
- Williams, B. (1995h). Voluntary acts and responsible agents. *Making sense of humanity and other philosophical papers, 1982–1993* (pp. 22–34). Cambridge University Press.
- Williams, B. (1995i). What has philosophy to learn from tort law? In D. G. Owen (Ed.), *The philosophical foundations of tort law* (pp. 487–498). Oxford University Press.
- Williams, B. (2002). *Truth and truthfulness: An essay in genealogy*. Princeton University Press.
- Williams, B. (2005a). *Descartes: The project of pure enquiry*. Routledge.
- Williams, B. (2005b). From freedom to liberty: The construction of a political value. In G. Hawthorne (Ed.), *In the beginning was the deed: Realism and moralism in political argument* (pp. 75–96). Princeton University Press.
- Williams, B. (2006a). The actus reus of Dr. Caligari. In A. W. Moore (Ed.), *Philosophy as a humanistic discipline* (pp. 97–108). Princeton University Press.
- Williams, B. (2006b). Moral responsibility and political freedom. In A. W. Moore (Ed.), *Philosophy as a humanistic discipline* (pp. 119–125). Princeton University Press.
- Williams, B. (2006c). Pagan justice and Christian love. In M. Burnyeat (Ed.), *The sense of the past: Essays in the history of philosophy* (pp. 71–82). Princeton University Press.
- Williams, B. (2011). *Ethics and the limits of philosophy*. Routledge.
- Wilson, G., & Shpall, S. (2012). Action. In E. N. Zalta (Ed.), *The Stanford encyclopedia of philosophy* (2016th ed.). Stanford University.
- Wittgenstein, L. (1958). *The Blue and Brown books: Preliminary studies for the philosophical investigations*. Blackwell.
- Yeager, D. B. (2006). *J.L. Austin and the law: Exculpation and the explication of responsibility*. Bucknell University Press.
- Zimmerman, M. J. (2002). Taking luck seriously. *The Journal of Philosophy*, 99(11), 553–576.