A Shelter from Luck:
Reconstructing the Morality System from the Ground Up

Matthieu Queloz

The "morality system," Bernard Williams concludes at the end of *Ethics and the Limits of Philosophy*, is "a deeply rooted and still powerful misconception of life" (2011, 218). It combines certain quite special conceptions of value, motivation, obligation, practical necessity, responsibility, voluntariness, blame, and guilt. But any attempt to characterize the morality system quickly runs the risk of degenerating into a laundry list of things that Williams happened to dislike. To see what holds the morality system together, we have to take a view of it that is sympathetic enough to recognize what this deeply rooted misconception is rooted *in*: what human needs and concerns does it answer to, and where do the ideas it draws on themselves come from? If Williams calls it a "system," it is because there are reasons for just those ideas to come together in just that way. Once we see the *point* of the system, we will be in a better position to see what is wrong with it, and why "we would be better off without it" (2011, 193).[1]

In order better to grasp the point of the system and why it combines the ideas it does in the way it does, I propose to reconstruct the morality system from the ground up: to ask not just why the system is as it is, but also why the ideas and practices it harnesses are there to be harnessed in the first place. This, it turns out, is a question Williams himself took very seriously. When Williams remarks that we would be better off without the system, he does not mean that we should completely jettison the conceptual material it draws on. Instead, his aim is the more constructive one of making "*some* sense of the ethical as opposed to throwing out the whole thing because we can't have an idealized version of it" (2009, 203). Throughout his work, one finds vindicatory explanations of the pre-morality-system versions of the

---

[1] Unless otherwise stated, all citations refer to works by Bernard Williams.

moral/non-moral distinction, the idea of obligation, the voluntary/involuntary distinction, the practice of blame, and other building blocks of the system. These vindications are explanatorily prior to, and importantly undergird, Williams's criticisms of the more particular forms that these ideas and practices assume within the system.

By juxtaposing vindication with critique, I hope to bring out the connections between them; in particular, I shall argue that understanding what these ideas do for us when they are not in the service of the system is just as important to leading us out of the system as the critique of that system, because this understanding provides a guiding sense of what we need these ideas to be, and offers us an alternative and more reflectively stable way of making sense of them in vindicatory terms.

It is crucial to Williams's critique that the morality system's "idealised version" of the ethical is not just a philosophers' fantasy that does little harm outside the seminar room, but "the outlook, or, incoherently, part of the outlook, of almost all of us" (2011, 194). It is *incoherently* part of our outlook because we also have, alongside the morality system's rarefied conceptions of such things as moral value, obligation, voluntariness, and blame, more everyday conceptions of these things, and a lot of the time, it is these more relaxed conceptions that we act on. Hence, when we talk simply of "our concepts" in the coarse-grained way that ignores finer distinctions between different conceptions of the same thing, there is an important sense in which "our concepts" are *not* those of the morality system. This explains how Williams can say, for instance, that "[w]e have fooled ourselves into believing that we have a more purified notion of moral responsibility than we have" (1999a, 163), or why he feels drawn to the distinction between what we *think* and what we *merely think that we think* (1993, 7, 91). But as this formulation itself brings out, the boundary between what we think and what we think that we think cannot ultimately be a sharp one: even what we merely think that we think will often have very real effects—not just on what else we think, but on how we end up living. The morality system may be an unrealisable vision, but it is not for all that unreal. It is the real problem of an unrealisable vision.

To understand how the morality system's conceptions and our more everyday conceptions can be seen as *conceptions of the same things* at all, and how they relate to each other, Williams approaches them in the light of a *tertium quid*: maximally *generic* conceptions of ideas and practices that are schematic, underdetermined, and

probably fictional, but that nevertheless present us with helpful models or proto-types of the conceptual practices we really do find in human societies. For Williams, the point of considering these prototypes mainly lies in identifying defeasible reasons to think that any human society—be it that of the Greeks or our own—will be bound to develop conceptual practices along these lines in *some* form, because these can be seen to answer to human needs of a basic and near-universal kind. To keep track of which of these three conceptions of a given conceptual practice is at issue, I shall use the subscript (GEN) to mark the *generic* conceptions that provide initial characterisations of conceptual practices and help us see the animating concerns behind them, but are too underspecified to be situated in space or time; (ORD) to mark the *ordinary* elaborations of these generic conceptions which we now in fact live by a lot of the time; and (MS) to mark the *morality system's* conceptions of these conceptual practices.[2]

My argument falls into three parts. The first part (§1) considers vindicatory explanations, in terms of highly generic and near universal needs, of what will turn out to be the four crucial building blocks of the morality system: the moral/non-moral distinction$_{(GEN)}$, the idea of obligation$_{(GEN)}$, the voluntary/involuntary distinction$_{(GEN)}$, and the practice of blame$_{(GEN)}$.[3] This part performs a double function: it explains why these conceptual practices are there to be harnessed by the system in the first place, and it offers us a way of making sense of them that is independent of the system. The second part (§2) is a vindicatory explanation, relative to the need

---

[2] How exactly the morality system's conceptions relate to our more ordinary and less demanding conceptions, and in what situations the former make themselves felt at the expense of the latter, are of course crucial questions for anyone seeking to understand the nature and scope of Williams's critique, but I must leave them for another occasion.

[3] A fuller treatment than I have room for here might add *guilt*, which Williams contrasts in particular with shame (1993, 1997a), and which he describes as "the characteristic first-personal reaction within the system" (2011, 197). The combination I repeatedly explore in this essay, of a vindication of the generic form of X with a critique of the refined form it takes within the morality system, can be found also in Williams's treatment of guilt. In its generic form, guilt helpfully "turns our attention to the victims of what we have wrongly done" (1993, 222). But this virtue is lost once the conception of guilt is elaborated into something more abstract in the morality system: "When the conception of guilt is refined beyond a certain point and forgets its primitive materials of anger and fear, guilt comes to be represented simply as the attitude of respect for an abstract law, and it then no longer has any special connection with victims" (1993, 222).

for ultimate fairness, of the way in which the morality system combines and refines these building blocks into the moral/non-moral distinction$_{(MS)}$, the idea of obligation$_{(MS)}$, the voluntary/involuntary distinction$_{(MS)}$, and the practice of blame$_{(MS)}$ in order to provide a shelter from luck. Reconstructing the system in light of this organizing ambition gives us a good grasp on why it has the shape it has, and what the different components of the system contribute. The third part (§3) is a critique of the resulting construction: I argue that the ultimate problem with the morality system is its frictionless purity. It robs valuable concepts of their grip on the kind of world we live in, and, by insisting on purity from contingency, threatens to issue in nihilism about value and scepticism about agency. To overcome these problems, it is not enough to accept that contingency and luck pervade human life. We also need to revise our understanding of what the facts of contingency and luck entail. In particular, we need to abandon the purist attitude that blinds us to alternative ways of making sense of human values and agency—alternatives that naturalistic but vindicatory explanations can provide.

1. Vindicatory Explanations of Four Building Blocks of the Morality System

1.1. The Moral/Non-Moral Distinction

Let us begin with what is arguably the most basic building block of the morality system: the moral/non-moral distinction$_{(GEN)}$. Critical as Williams may be of the particular form which the distinction between the moral and the non-moral takes within the morality system, he still has a vindicatory story to tell about our need to draw *some* distinction along those lines. One of the aims of his 1972 book *Morality*, Williams declared in the preface he added in 1993, was the "placing of morality in relation to other ethical considerations and to the rest of life" (2001, xiv). The book achieves this, notably, by examining "what the distinction between the 'moral' and the 'nonmoral' is supposed to do for us" (2001, xiii). Already in his first book, therefore, Williams is concerned to do what he more explicitly aims to do in his last book, *Truth and Truthfulness* (2002): to situate philosophically puzzling concepts and distinctions in human affairs by relating them "to other reasons for action that human beings use, and generally to their desires, needs and projects" (2001, xiii).

As Williams makes clear in that preface, his 1972 book did not yet observe the distinction he later came to draw between the "moral" in the narrow sense in which

that idea is understood within the morality system and the broader notion of the "ethical." Accordingly, his 1972 enquiry still concerns the moral/non-moral distinction(GEN), of which the moral/non-moral distinction(MS) is a particular socio-historical elaboration.

Williams suggests that the moral/non-moral distinction(GEN) is one we should expect to find at some level in any human society. As he emphasises in his discussion of the amoralist, it may be possible for an individual to live outside the ethical life, but no human community can get by without some kind of morality, i.e. without some minimal ethical consciousness which stakes claims against self-interest.[4] To stake claims against self-interest, Williams contends in "Egoism and Altruism," is "one basic and universal function of morality" (1973b, 250).[5] Any morality, in order to count as a morality at all, needs to involve some distinction between actions which only minister to the interests of the agent at the expense of others and actions which take the interests of others into account. "If some such distinction is not made," Williams insists, "there are no moral considerations at all" (2001, 66). In its most primitive form, the concept of the *moral* marks a distinction between selfish actions and other-regarding actions and selects the latter for approval.

Of course, this primitive distinction between two classes of actions is still too primitive; if we are to make sense of anything like *our* concept of the moral, we need to understand why it involves discriminating not just between different kinds of action, but also between different kinds of *motivation*. The explanation that Williams gives turns on the idea that actions can be other-regarding while springing from self-regarding motives. He gives the example of "a self-interested business man who writes a cheque to famine relief" (2001, 66), but whose concern is for his own reputation rather than for the relief of famine. "What," Williams asks, "is the point and content of saying that we do not *morally* approve of the self-interested donor to charity, or that, though he does a good thing, he does not act morally?" (2001, 67). As we will see, the idea that to act *morally* is to act *from moral motives* will acquire a central

---

[4] See Williams (2011, 32, 51).

[5] On this usage of the term "morality," altruism is necessary for morality, where altruism is construed broadly as "a general disposition to regard the interests of others, merely as such, as making some claim on one, and, in particular, as implying the possibility of limiting one's own projects" (1973b, 250).

importance within the morality system; but on Williams's view, we have antecedent reasons to refine the moral/non-moral distinction far enough to discriminate not just between actions, but also between the motives from which they spring.

Williams invites us to consider three types of actions: (i) self-regarding actions, (ii) other-regarding actions done from self-regarding motives, and (iii) other-regarding actions done from other-regarding motives. Williams highlights that there are good reasons for a community to encourage (ii) over (i): an other-regarding action done from self-regarding motives is still better than a selfish action—the self-interested donor will, after all, help relieve famine, and this is surely "better than that another combined cocktail cabinet and TV set should be bought" (2001, 66).[6] What is more, approval need not be limited to the self-interested donor's *action*: we can also approve *of him* insofar as he has done something which it is a good thing to have done.

However, Williams also thinks that there is "a very good point" (2001, 67) in withholding *moral* approval in the case of the self-interested donor, i.e. in reserving the kind of approval that comes with one's action being considered a *moral* action for cases of type (iii). To draw a contrast between (ii) and (iii) is to contrast motivations which will only yield other-regarding acts when these happen to align with the agent's self-interest with motivations that are *steadier* because they are general dispositions to do things of the other-regarding sort: motivations grounded in principle, for example, or motivations grounded in sympathy with others. Such motivations will more reliably give rise to other-regarding actions than motivations which are conditional on those actions aligning with self-interest. "This must surely," Williams remarks, "have something to do with the *point* of selecting certain motives for *moral* approbation: we are concerned to have people who have a general tendency to be prepared to consider other people's interests on the same footing as their own" (2001, 68).[7]

The point of having a concept of the moral that ties moral action to moral motive,

---

[6] In the fifties, the "combined cocktail cabinet and TV set" really was a thing. Williams's suggestion that this would have been a particularly frivolous acquisition is vindicated by the fact that it now sounds to us rather like Boris Vian's fictional "pianocktail."

[7] There is a parallel here to Philip Kitcher's claim, in *The Ethical Project* (2011), that the most basic point of morality is to remedy "altruism failures."

then, is to cultivate in people a general tendency to engage in other-regarding behaviour by selecting for moral approval "general dispositions to do things of the non-self-interested sort" (2001, 69–70).[8] So far, this is a vindicatory story. We get a vindicatory explanation, first, of why any human community would need to draw some kind of distinction between moral and non-moral actions, and then of why this distinction would need to be focused further to make a moral action one that stems from a certain kind of motive.

1.2 The Idea of Obligation

Another crucial source material for the morality system that Williams aims to achieve an independent grip on is the idea of obligation(GEN). Why do we have it in the first place? What does it do for us? In *Ethics and the Limits of Philosophy* and in his essay "The Primacy of Dispositions," Williams goes some way towards offering an explanation that might "help us to understand the point and value of living a life in which obligations counted as ethical reasons" (2006e, 73).

The idea of obligation(GEN) (along with its correlate, the idea that those towards whom one has an obligation have a *right*)[9] is grounded in the basic interest of human beings everywhere in being able to *rely* on certain things—that they will not be killed, assaulted, or arbitrarily expropriated, for example (2011, 205). The idea of obligation (GEN) works to secure reliability about such things of importance by helping to create "a state of affairs in which people can reasonably expect others to behave in some ways and not in others" (2011, 208). In particular, it works to ensure that considerations of importance are given *high deliberative priority*. This, according to Williams, is the most basic point of the idea of obligation(GEN): it gives important considerations high deliberative priority. A consideration enjoys such priority for us if, first, it appears in our deliberations, and second, it is given heavy weighting against other considerations (2011, 203). The concept of obligation(GEN) is thus like the special email format which ensures that important emails are flagged as "high priority" when they appear in the recipient's inbox: it provides a format for ethical

---

[8] To say this is not to deny that having such a concept also serves other functions—Williams is explicit about this even in *Morality* (2001, 69), and, of course, he takes the concept of obligation(MS) to perform different and less benign functions.

[9] See Williams (2011, 206).

considerations that lends them particular prominence and weight in people's deliberations.

Using the idea of obligation, we can try to make sure that considerations of basic and standing importance are reflected in "settled and permanent pattern[s] of deliberative priorities" (2011, 206). But obligations can also be more context-sensitive. If the person next to me suffers a stroke, a "general ethical recognition of people's vital interests" becomes "focused into a deliberative priority by immediacy" (2011, 206). Immediacy *to me* generates an obligation *for me* to help. Even more conditional are obligations generated by promises. The institution of promising "operates to provide portable reliability," as Williams puts it, "offering a formula that will confer high deliberative priority on what might otherwise not receive it" (2011, 207–208).

This account is reminiscent of Hume's genealogy of the respect for property and the institution of promising, and indeed, Williams later noted that his account "is broadly in the spirit of Hume" (1995g, 205).[10] It presents obligations as one notable way of "securing the protection of important interests" (1995g, 205), but the kind of importance in terms of which the notion of obligation is explained is of an everyday, naturalistically unmysterious sort: it is the importance of satisfying elementary needs.

In light of this, it is not altogether surprising that the morality system should have grown around the notion of obligation rather than around some other category of ethical thought. If Williams is right, the notion of obligation is a device which originally serves to acknowledge and reflect, at the level of deliberation, the overriding urgency and demandingness of our most basic needs and our needs in situations of emergency. Understanding these practical origins helps us understand why the device is so demanding—as Williams puts it, it is "just because the needs involved are so elementary that the psychological mechanisms designed to meet those needs are demanding" (1995g, 205). Yet is also "because those mechanisms are demanding that the theory which grows around them becomes so dense and oppressive" (1995g, 205). Taken beyond its proper remit, the device of obligation can soon seem absurdly overpowered, rather like a Roman dictator who retains his emergency powers beyond the fulfilment of his mandate.

---

[10] See Hume (2000, 3.2.2 and 3.2.5) and Queloz (Manuscript, chap. 3).

1.3 The Voluntary/Involuntary Distinction

The third crucial building block for the morality system is the voluntary/involuntary distinction(GEN). That distinction, Williams thinks, can be constructed already out of distinctions that human beings everywhere are bound to find worth having. "All conceptions of responsibility make some discriminations" (1993, 66) between what is voluntary and what is not, on his view. At the same time, he is adamant that "no conception of responsibility confines response entirely to the voluntary" (1993, 66). But he nonetheless maintains that the voluntary/involuntary distinction, and the concomitant notion of "the will" in a correspondingly unambitious sense, are universally worth having.

His idea is that the need for the voluntary/involuntary distinction(GEN) grows out of the need for some practice of recognizing responsibility(GEN). Already in his essay for the BBC's *Third Programme*, published in 1963, Williams observes that this is a practice which the Greeks shared with us in some form (1963, 1–2). Like us, they recognized that to be responsible for a certain state of affairs is not *just* a matter of being the cause of it through some movement of one's body. To determine whether someone really carries *responsibility* for that state of affairs in the full sense which makes that person a proper subject of blame, we want to know more about the bodily movement. Was it just a nervous twitch, or did they really *act*? Did they *intend* to bring about that state of affairs? And what *state of mind* were they in when they did so? Williams elaborates on this generic conception of responsibility in *Shame and Necessity*, where he spells out four "basic elements of any conception of responsibility" (1993, 55):

> *Cause*: the idea that someone brought about a bad state of affairs in virtue of what they did;
>
> *Intention*: the idea that they intended that state of affairs;
>
> *State*: the idea that they were in a normal state of mind when they brought it about (i.e. not sleepwalking or subject to extreme incident passions);
>
> *Response*: the idea that this calls for some response on their part, that they need to make up for it.

Williams's thought is that out of these four basic elements—easy to remember be-

cause they form the acrostic *CISR*—a great many different conceptions of responsibility can and have been constructed by interpreting the elements in different ways and varying the emphasis between them. These are "universal materials" (1993, 56), because the need for them follows "simply from universal banalities" (1993, 55):

> Everywhere, human beings act, and their actions cause things to happen, and sometimes they intend those things, and sometimes they do not; everywhere, what is brought about is sometimes to be regretted or deplored, by the agent or by others who suffer from it or by both; and when that is so, there may be a demand for some response from that agent, a demand made by himself, by others, or by both. (1993, 55)

It may be plausible enough that the ideas of *Cause* and *Response* are universal, but why should we expect the ideas of *Intention* and *State* to be universal? Williams's answer, in a nutshell, is that we are bound to be interested in drawing some distinction between what is intended and what is not, and what is done in a normal state of mind and what is not, because these distinctions are crucial to understanding how an action relates to the plans and character of the agent. They make a great difference to our relations to other people.

Williams gives an example from the end of the *Odyssey*, where Odysseus and Telemachus confront Penelope's suitors and find, to their great alarm, that the suitors are handing out the weapons that Telemachus was supposed to have hidden away in a storeroom. Odysseus angrily wonders who opened the storeroom, and Telemachus explains that it was his mistake, and that no one else is to blame—he left the door of the storeroom ajar, and one of the suitors must have been a better observer than he was (1993, 50). Telemachus is clearly drawing a distinction 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, Williams contends, shows that although Homer had no direct equivalent for the *word* "intention," he nevertheless 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 (1993, 50–51). For Williams, it is no surprise that they draw such distinctions: "it must be a possible question how the intentions and actions of an agent at a given time fit in with, or fail to fit in with, his intentions and actions at other times," he writes, because "[u]nder any social circumstances at all, that is a question for other people who have to live

with him" (1993, 56). Being sensitive to which aspects of an action are intentional helps us understand what kind of action it is, and what else to expect from someone who, in that situation, intends those things. If Telemachus had intentionally left the door ajar, this would have disquieting implications for Odysseus, suggesting that Telemachus was not, after all, on his side. Here, the question whether Telemachus *meant* to do what he did is a matter of life and death.

Similarly, we are bound to care about whether actions are done from a normal state of mind or not, because this is crucial to figuring out what to expect from people in the future: if someone acts intentionally but in a strange state of mind, we know that their actions probably do not stand in a very regular relation to their overall plans and character. Williams gives the example of Agamemnon, who took Briseis from Achilles, and did so intentionally, but was in a strange state of mind when he did so: the gods cast *ate* on his wits, so that he was in a state of delusion or blind madness (1993, 52). (The example also brings out that ideas of what normal and abnormal states of mind *are* can vary.) Being sensitive to the state of mind from which people act helps us see how their intention and action on one occasion fits into the broader pattern of their intentions and actions on other occasions. It helps us separate the exceptional from the expectable. Like the capacity to separate the intentional from the unintentional, this is a capacity that human beings living together are bound to have an interest in possessing.

This brings us to what, for our purposes, is Williams's key point, namely that if the notions of *Cause*, *Intention*, *State*, and *Response* are available, one already has all the material necessary to construct the notion of the voluntary(GEN). It earns its keep in virtue of the human need to recognise responsibility for certain actions and to understand the place of intentions and actions in people's plans and characters.[11] The notion of the voluntary(GEN) picks out all those actions we are left with once we have filtered out things done unintentionally or in an abnormal state of mind: "a certain thing is done voluntarily if (very roughly) it is an intentional aspect of an action done in a normal state of mind" (1993, 66). And it is a similar notion of the

---

[11] Williams puts it even more strongly in a footnote to "Moral Luck: A Postscript": "the idea of the voluntary … is inherent in the concept of action" (1995e, 247 n. 4). See also Williams (2002, 45, 1995c, 1995f, 1999a; Magee and Williams 1971).

voluntary that Williams sees at work today. Speaking about our present-day version of the voluntary/involuntary distinction$_{(ORD)}$ in a later essay, he writes: "'A does X voluntarily' is equivalent to 'A does X intentionally in a normal state of mind'" (2006a, 120).[12]

These notions of the voluntary—be it voluntary$_{(GEN)}$ or voluntary$_{(ORD)}$—may strike one as superficial: if we push beyond a certain point such questions as what exactly someone intends, what makes it true that they intend it, and whether they intended to become the kind of person who can intend such a thing, such a notion of the voluntary gives out. But for Williams, its superficiality is precisely what makes it worth having: it is "essentially superficial" (1993, 67), and "if voluntariness is to do its work such questions cannot be pressed beyond a certain point" (2006a, 124). It is "an essentially superficial notion, which works on condition that one does not try to deepen it" (1997b, 495). A useful notion of the voluntary is one that helps us capture such obviously important differences as that between intentionally turning on the stove and somnambulism. Distinctions at this superficial level do nothing to settle the problem of free will (nor, indeed, do they generate that problem in the first place).[13] But it is by doing work at this superficial level that the notion of the voluntary helps us to live. It is (in Nietzsche's phrase which Williams quotes more than once) superficial out of profundity.[14]

1.4 The Practice of Blame

The last building block of the morality system I want to examine, which also grows out of the practice of recognizing responsibility, is the practice of blame$_{(GEN)}$. Here as

---

[12] Another formula he uses is: "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" (1995i, 25).

[13] See Williams (1993, 67–68): "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 … there is a problem of free will only for those who think that the notion of the voluntary can be metaphysically deepened."

[14] Nietzsche himself uses the phrase more than once—first in *The Gay Science* (2001, Preface, §4) and then again in *Nietzsche contra Wagner* (2005, Epilogue, §2). For further discussion of the superficiality of the notion of the voluntary, see Williams (1993, 67–68, 1995b, 127–128, 1995e, 243, 1997b, 495–496, 2006a, 124–125).

well, Williams's critique of blame$_{(MS)}$ is rooted in a more charitable account of blame$_{(GEN)}$, an account which presents blame$_{(GEN)}$ as a valuable, if peculiar, "instrument of social control" (1995c, 15). It is an instrument of social control because it helps sustain human communities by inducting new individuals (including notably children) into a shared ethical sensibility, and by helping to secure some degree of realignment with that sensibility where individuals problematically deviate from it. It is a peculiar instrument, however, because as long as the participants in the practice of blame think of it in purely instrumental terms, as justified only by its *efficacy* as a tool of alignment, blame will fail to produce the desired effects—it will tend to produce resentment rather than remorse.

> Blame that is perceived as unjust often fails to have the desired results, and merely generates resentment. This shows that the idea of blame's justification is not the same as the idea of its efficacy. When a recipient thinks that blame is unjustified, the content of his thought cannot be that the blame will be ineffective. This does not show that the purpose of blame may not in fact lie in the modification of behaviour; it means only that if this is true, it cannot be obvious to those who are effectively blamed. (Williams 1995c, 15)

The practice of blame can be efficacious only insofar as it is understood by participants to be more than just a regulative device, because only then will recipients of blame be suitably moved by the normative demands that blame expresses. Consequently, no account of blame that bases its justification merely on its efficacy or functionality can be adequate, because "it collides with one of the most obvious facts about blame, that in many cases it is effective only if the recipient thinks that it is justified" (1995c, 15).

At first pass, Williams's conclusion that blame cannot wear its function on its sleeve seems subversive rather than vindicatory, thereby calling into doubt my claim that Williams had a vindicatory account of blame in its generic form. But there are two ways of hearing the claim that blame cannot wear its function on its sleeve. On the first interpretation, which, admittedly, is strongly suggested by some passages in Williams's writings from the 1980s, he is indeed critical even of blame$_{(GEN)}$: he is saying that there is bound to be a *tension* between the non-instrumentalist spirit involved in the practice of blame and the instrumentalist spirit involved in grasping its point. If the practice of blame is to be stable despite this tension, the argument then goes, some way must be found of driving a wedge between what justifies blame from the point of view of the participants, and what justifies blame from a detached,

reflective point of view. But this yields a "Government House" account of blame akin to what Williams dubbed "Government House" utilitarianism (1995h, 166). It "most naturally fits a situation in which those who understand the justification, and those whose behaviour is being modified, are not the same people" (1995c, 15), and therefore invites the same concerns about social transparency that utilitarianism does. And even if one instead tries to compartmentalise the individual consciousness using some distinction between theory and practice, and assumes that participants are aware of blame's efficacy only off-duty, in the "cool hour" of reflection, this distinction between theory and practice possesses no real "saving power" (1995h, 165). The tension will manifest itself even in a compartmentalised mind, and the practice of blame must unravel under reflection.

But there is a second way of hearing the claim that blame cannot wear its function on its sleeve, one which does not commit us to there being a tension in the first place. Taking a leaf out of *Truth and Truthfulness*, we can see blame as a practice exhibiting what I call *self-effacing* functionality:[15] the practice of blame is functional, but only insofar as and because it is sustained by motives and reasons that are autonomous, i.e. not conditional on the practice's functionality in each particular case. The practice must outrun its functionality in order to be functional. As a result, the functionality of blame is inconspicuous or effaced, i.e. it is not the primary consideration for participants as they engage in the practice, but *for benign functional reasons*: we reap the benefits of blame only if we are bloody-minded about it rather than benefit-minded.[16] In this sense, the functionality of blame is *self-effacing*.

This account of blame also explains why blame's functionality is not obvious to participants, because that very functionality requires that their motivations and reasons focus on something other than the practice's functionality, and hence blame's functionality will efface itself in favour of those other motivations and reasons; but it crucially differs from the first account in that it does not present the functionality of blame as *necessarily* effaced: participants can become fully conscious of it without

---

[15] See Queloz (2018).

[16] A phrase Williams uses to describe similar functional dynamics in his account of truthfulness (2002, 59).

the insight into the functionality having a destabilising effect on the practice, because not only the practice of blame, but also the participant's practice of thinking of it non-instrumental terms are vindicated, and are vindicated in terms that (unlike the sort of vindication envisaged by indirect utilitarianism) do not undermine the authority of the non-instrumental reasons for blame.[17] Indeed, in *Truth and Truthfulness*, the insight into the functionality of thinking in non-functional terms is explicitly offered as something by which to *strengthen* our confidence in the practice at issue.

On this reading of Williams on blame, we get an explanation of why we go in for blame that is in the first instance vindicatory rather than subversive—one, moreover, that does not just vindicate blame as a salient solution to the problem of securing the alignment of ethical sensibilities required for humans to live together, but that also specifically vindicates the peculiar self-effacingly functional structure of blame: blame is needed as a device of ethical alignment, but it also needs to be *more* than a device; in particular, it needs to be *justified* by more than its functionality in order to be functional. Indeed, Williams's account of the point of blame might be taken to suggest that we who live in heterogenous liberal societies have a *special* need for blame, since the more diverse societies are, and the more they deny themselves more draconian ways of securing alignment in behaviour, the more there is a need for blame.[18]

For Williams, then, the four building blocks of the morality system I have focused on—the moral/non-moral distinction(GEN), the idea of obligation(GEN), the voluntary/involuntary distinction(GEN), and the practice of blame(GEN)—are not the invention of the morality system. They have deeper roots in more generic needs.

2. A Vindicatory Explanation of the Morality System

With these vindicatory explanations of the building blocks of the system in place,

---

[17] This is a highly condensed summary of an argument elaborated in Queloz (Manuscript).

[18] Nor does the fact that blame sometimes over-stretches the idea that one had reason to act otherwise seem to count against it on Williams's reckoning, since even the fiction that one had reason to act otherwise has a valuable tendency to instil in the blamed just the sensitivity to reasons that it pretends they already possess (1995d, 41–44). On this "proleptic" function of blame, see Miranda Fricker's Williams-inspired "What's the Point of Blame? A Paradigm Based Explanation" (2016).

we can now turn to the system itself. To what end does it incorporate and reshape just these building blocks in the way it does? What needs and concerns explain the "particular development of the ethical" (2011, 7) that is the morality system, and what is its point?[19]

Williams gives his most pointed answer to this question in "Moral Luck: A Postscript," where he writes:

> [T]he point of this conception of morality is, in part, to provide a shelter against luck, one realm of value (indeed, of supreme value) that is defended against contingency. (1995e, 241)

Our need for a shelter from luck grows, Williams suggests, out of a longing for "ultimate justice" (2011, 43) or "fairness" (1995f, 75). In the face of the fact that "most advantages and admired characteristics are distributed in ways that, if not unjust, are at any rate not just, and some people are simply luckier than others," the morality system expresses "the ideal that human existence can be ultimately just," because it offers a special kind of value, *moral* value, that outshines every other kind of value and at the same time "transcends luck" (2011, 217).[20]

On Williams's view, then, the need or concern that has to be factored in to get from the moral/non-moral distinction(GEN), the idea of obligation(GEN), the voluntary/involuntary distinction(GEN), and the practice of blame(GEN) to the moral/non-moral distinction(MS), the idea of obligation(MS), the voluntary/involuntary distinction(MS), and the practice of blame(MS) is the *longing for ultimate justice or fairness*, which these notions serve to deliver by providing *a shelter from luck or contingency*. Is that longing for ultimate justice or fairness a more socio-historically local need than the generic human needs we considered above? Williams's position on this is difficult to pin down. On the one hand, he seems to view the morality system as a broadly speaking "modern" phenomenon that is historically connected to Platonism

---

[19] A more general, but related, question is what drives the systematization of ethical thought that ultimately issues in ethical theory in the first place. See Cueni and Queloz (Forthcoming) for a reconstruction of Williams's answer to this question.

[20] Many have since highlighted the importance of justice or fairness as a motivation for the immunisation of morality against luck; see Latus (2000, 166), Levy (2011, 9–10), Otsuka (2009, 374–75), Sher (2005, 180), Statman (2005, 425), and Walker (1991, 16). For a critical discussion of various fairness-based arguments in favour of luck-free morality, see Hartman (2016, 2017).

and Christianity and that finds its purest expression in the moral philosophy of Kant.[21] On the other hand, he seems sympathetic to the idea that the sense of fairness and the resentment of unfairness have deep naturalistic roots in the social character of our species, and might even be innate.[22]

But as Williams also wrote, what needs and desires we have is a function of what we deem possible (1973a, 147), and perhaps the thought is that the longing for fairness is old, but the *sense that ultimate fairness is possible* is not: that, very roughly speaking, had to wait for Plato, who, by developing Pythagorean ideas,[23] provided the required dualism of soul and body, and in particular the idea of a "featureless moral self" (1993, 160). When Plato had Socrates declare that "*the good man cannot be harmed*, since the only thing that could touch him would be something that could touch the good state of his soul" (2011, 39), he was articulating the animating idea of the morality system. It was also Plato who imposed a stark division between "rational concerns that aim at the good, and mere desire" (1993, 42), thereby providing the strategy of *ethicizing psychology*, as Williams calls it—the strategy of fitting psychological ideas to moral demands instead of trying to fit moral ideas into an independent understanding of human psychology.[24] So while the concerns driving the construction of a shelter from luck may be much older, the construction itself had to wait for suitable material to come together. Only then could the ambition to "make the world safe for well-disposed people" (2006f, 59) in the way that the morality system proposes gain a foothold.[25] Refracted through Platonic ideas, the primitive need for fairness is elaborated and focused into a need for *ultimate* fairness. In this "strong form," Williams notes, the need to resist luck is one of the "idiosyncrasies" of the "local species of the ethical" (1995e, 242) that is the morality system.

Once we see the morality system as organised by a concern to serve the need for

---

[21] See Williams (2014, 86).

[22] See Williams (1999b, 248).

[23] See Williams (2006d, 16).

[24] In Williams's own terms, to ethicize psychology is "to provide a psychology that gets its significance from ethical categories" (1993, 43) or to define "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, 160). See also Williams (2006b, 78).

[25] A phrase Williams used in the referenced passage to describe "the tireless aim of moral philosophy."

ultimate justice by providing a shelter from luck, we can reverse-engineer the distinctive contributions of its various components and offer an explanation of their elaboration into the moral/non-moral distinction(MS), the idea of obligation(MS), the voluntary/involuntary distinction(MS), and the practice of blame(MS). To the extent that these elaborations are exhibited as contributing to the satisfaction of the need for ultimate justice, that explanation will in the first instance be vindicatory. To this end, our guiding question must be: *How does one construct a shelter from luck?*

First, one needs a special kind of value that is not "merely a consolation prize you get if you are not in worldly terms happy or talented or good-humoured or loved," but the "supreme" form of value—it "has to be what ultimately matters" (2011, 217) if it is to eclipse any lack of luck in other respects. This is where the *moral/non-moral* distinction(GEN) comes in. In its generic form, Williams agrees with Hume, the distinction is not very sharp.[26] But it can be elaborated and exaggerated into a stark distinction to provide a special kind of value—*moral*(MS) value—whose importance can then be dialled up to the point where it drowns out any other kind of value. The supremacy of moral(MS) value importantly contributes to shutting out luck by ensuring that misfortunes in dimensions other than the moral do not count. The most effective way of achieving this is for moral(MS) value to be supreme not just in the sense of carrying more weight than other kinds of value, but of forbidding comparison altogether. It is not that moral(MS) value ends up *outweighing* other kinds of value—there is not even a competition. This is what Williams means when he notes the tendency of the morality system to *close in on itself*, so that it comes to seem an "indecent misunderstanding" (2011, 217) to ask, as Nietzsche does, what the value of that system is.[27] The morality system's "purism and its self-sufficiency mean that it is structured not to hear any considerations that might limit its own" (1995g, 204). From the point of view of the system, nothing outside the system really matters.

Second, one needs to ensure that the point of view of the system, and the demands it makes on us, are truly inescapable. This means that the demands raised by

---

[26] See Williams (1995c, 20 n. 12) and Hume (1998, Appendix IV).

[27] In the *Genealogy*, Nietzsche asks after "the *value* of morality," urging that "we need a *critique* of moral values, *for once the value of these values must itself be called into question*" (1998, Preface, §§5–6). Williams echoes Nietzsche when he notes that "the principal aim of all moral philosophy" is that of "truthfully understanding what our ethical values are and how they are related to our psychology, and making, in the light of that understanding, a valuation of those values" (1995a, 578).

the system have to combine two aspects: *ubiquity*, so that there is no domain in which the demands of the system do not arise; and *stringency*, so that these demands are forceful enough not just to get a hearing, but to take precedence over other demands. This is where the notion of *obligation*$_{(GEN)}$ comes in as the ideal format in which to couch moral thought. For the reasons we considered, obligations are stringently demanding, and are designed to intrude into deliberation and impose themselves at the top of the priority list. Moreover, as Williams notes, "[i]f obligation is allowed to structure ethical thought, there are several natural ways in which it can come to dominate life altogether" (2011, 202). If only an obligation can overrule an obligation, the felt need to resist a given obligation will invite one to look for ways in which the need to resist it can itself be rationalized as expressing a general obligation; and the more this happens, the more general obligations multiply, so that they end up providing "work for idle hands" (2011, 202) across all wakes of life. The notion of moral obligation$_{(MS)}$ is thus perfectly suited to the task of ensuring that the system will be truly inescapable. The proliferation of moral obligations$_{(MS)}$ means that there is always something that one is under an obligation to do, so that there is no sphere of life that the morality system fails to reach into; and the stringency of moral obligation$_{(MS)}$ ensures that what the morality system demands is what one really *must* do. The notion of obligation$_{(GEN)}$, once elaborated into the notion of obligation$_{(MS)}$, allows the demands of the morality system to become *categorical*—a Kantian term, which, as Williams notes (2011, 198), usefully combines the two aspects of ubiquity and stringency.

Our construction thus far already looks pretty well defended against contingency. Insofar as agents manage to stand in the right relation to their moral obligations, they will be living well in the only respect that ultimately matters. Insofar as they manage to enter the shelter, therefore, they will be safe from luck. And everywhere, the shelter is there to be entered, because everywhere, morality makes a claim on one—and not just a claim, but *the* claim: everywhere, what morality demands is what one really must do. We have a robust shelter from luck which ubiquitously invites us in and promises to shut out the raging elements.

But though ubiquitous and contingency-proof in itself, our shelter from luck does not yet effectively serve people's need to escape contingency, because it potentially still suffers from what might be called an *unequal access* problem: some people may, for contingent reasons, find it easier than others to align their lives with the demands

of the morality system. To eliminate contingency even here, we need to regulate entry to the shelter in terms that guarantee equal access. Clearly, for example, moral value must not be tied up with the consequences of actions, because this would again render the agent vulnerable to incident luck. Taking a sufficiently long-term view of the consequences of an action can effectively leave us clueless about its moral value—as a Medieval proverb has it, when the flung stone leaves the hand, it belongs to the devil.[28] To truly discount luck, moral value therefore needs to retreat into the agent, to lie "in trying rather than succeeding, since success depends partly on luck" (2011, 217). Harnessing the emphasis on moral motivation which we noted already in the moral/non-moral distinction$_{(GEN)}$, the system is thus driven to focus exclusively on the agent's moral intentions and motives.[29]

Yet if the basis on which we allocate moral worth is to be ultimately just, we still need to eliminate various contingencies *within* the agent as well, because, as Williams points out, the "capacity to try," or to act from moral motives, "is itself a matter of luck" (2011, 217). Various contingencies at the level of natural inclinations or endowments, socialisation, education, and other biographical and historical circumstances may make it easier for some agents to develop the right kind of motivation. Moral motivation must therefore itself be understood in terms that insist on purity from contingency. It must not be conditional on the agent's contingent desires or motives. It must be a form of motivation that the agent has *anyway already*—for instance, in virtue of being a rational agent.

The requirement that makes itself felt here is that in order to guarantee equal access to the shelter from luck, the morality system must base itself solely on what any agent has complete control over no matter their circumstances. This is where our third component, the *voluntary/involuntary* distinction$_{(GEN)}$, comes in. Even in its generic form, it already does some of the work required by separating out what the agent unintentionally did or did in an abnormal state of mind. But the "search

---

[28] On the problem of cluelessness, see Greaves (2016).

[29] Some attempts to render utilitarianism actionable issue in a notably similar structure: by enjoining the agent to maximize *expected* rather than *actual* utility, utilitarianism condenses into an obligation to do whatever has the highest expected utility to the best of the agent's knowledge, and although how much actual utility this produces is to a great degree a matter of luck, agents escape blame as long as they maximize expected utility (at least most of the time—see Monton (2019) for a discussion of exceptional cases).

for an intrinsically just conception of responsibility" (1993, 95) leads one to push responsibility even further back, to a *purified* form of trying that is not conditioned by any disposition or desire, or any other trait that the agent contingently has: what Williams calls "utter voluntariness" (2011, 218) or, following Kant, "the unconditioned will" (1981, 20). To be truly unconditioned, of course, this purified conception of "the will" may not be in any way empirically determined by what the agent contingently is. So the locus of the will in this demanding form cannot be the socially situated and contingently constructed self. It has to be the *featureless moral self* that lies beyond all determination by contingent empirical circumstances (in Kantian terms: the noumenal or transcendental self). The intrinsically just and ultimately fair basis on which moral responsibility(MS) is allocated in the morality system is thus the unconditioned will of the characterless self.

It should be acknowledged in passing that this idea of the unconditioned will, which has a rich history in the theory of action,[30] also has other roots, as Williams himself acknowledges in various places. The morality system is in that sense "overdetermined" (1995g, 204). Williams notes, for example, that the "phenomenology of bodily movement and the notion of trying" (1999a, 149) already invite, via the observation that one can will a movement without that movement actually ensuing, the distinction between the self *qua* locus of action and the self *qua* locus of the will. He also remarks that the notion of the will as something free from empirical determination comes in because we want there to be something over we which we have complete control, and we want that because we feel the need for "*real authorship* of our actions" (1999a, 149). A further driving force is the powerful feeling of resentment we feel when others wrong us. Williams here takes a leaf out of Nietzsche's *Genealogy*, which describes the idea of a featureless moral agent—an agent lying beyond all determination by circumstance who can will to actualize his contingent dispositions or not—as being motivated by the felt need to blame, not just the nature of things in general, but those who wronged us in particular: by postulating the idea that agents are free to transcend their nature, the wronged ones "gain for themselves the right to hold the bird of prey *accountable* for being a bird of prey," because it allows them to say that "*the strong one is free* to be weak, and the bird of prey to be

---

[30] For overviews of that history, see, e.g., Glock (1996, 'will'), Hyman (2011), Candlish and Damnjanovic (2013), and Queloz (2017).

a lamb" (Nietzsche 1998, I, §13). Connected to this is the human tendency to indulge in what Williams calls a "special fantasy of retrospective prevention" (1995f, 73), where the victim fantasizes about replacing the wrongdoer's action with an acknowledgement of the victim. This fantasy again motivates thinking of the agent in isolation from the network of circumstances, as an autonomous entity capable of willing to act otherwise than the agent in fact did.

Lastly, the morality system must also allocate moral *blame* in a way that shields the agent from luck. This is easily achieved at this point, since moral blame$_{(MS)}$ only needs to track utter voluntariness to make sure that the agent is blamed only "on the ultimately fair basis of the agent's own contribution" (2011, 216). This is what Williams calls the "purified conception of blame" (1995f, 72). By tying blame to utter voluntariness, the morality system ensures that agents are blamed "for no more and no less than what is in [their] power" (1995f, 72). In effect, given the moral demands on the will$_{(MS)}$ take the form of obligations$_{(MS)}$, the purified conception of blame ties blame$_{(MS)}$ to the *purely voluntary breaking of obligations*. Because the morality system focuses on blame$_{(MS)}$ at the expense of other reactive attitudes and links blame$_{(MS)}$ to the purely voluntary breaking of obligations, "[t]he thought *I did it* has no special significance" within the resulting picture of the ethical life; the only question is "whether I voluntarily did what I ought to have done" (2011, 196). This leads to a blinkered disregard for what Williams insists is an important dimension of ethical experience, namely "the distinction simply between what one has done and what one has not done" (2011, 196).[31]

The purified conception of blame comfortingly shields one from two kinds of blameworthiness that would otherwise render one vulnerable to luck. On the one hand, it shields one from being blamed for what one does involuntarily. (This is what George Sher aptly calls the "Searchlight View" (2009, chap. 1) of responsibility, on which agents are responsible only for those features and results of their acts of which they are *aware* when they perform the acts.) On the other hand, it shields one from incurring blame when one does something as the lesser of two evils. These fateful choices between wrong and wrong—the stuff of tragedy—lose their sting in the morality system, because if blameworthiness is tied to broken obligations, and if *ought*

---

[31] This, I take it, is one of the main messages of "Moral Luck" (1981). See also Williams (2011, 43–44).

implies *can*—one can only be under an obligation to do what one can do—one is not blameworthy when one does something as the lesser of two evils.[32] There might have been what W. D. Ross (1930) calls a *prima facie* obligation not to do what one ended up doing, but this obligation was eventually defeated by the consideration that the alternative would have been worse, and so there was no *real* obligation not to do what one did.

Once the practice of blame(MS) is appropriately purified to be sensitive only to purely voluntary acts, our moral agents longing for ultimate justice are finally home and dry. The only thing that ultimately matters—moral(MS) value—is now completely within their control, for it depends only on whether they choose, from motives they all equally have anyway, to align their unconditioned will with their categorical obligations.

3. Critique of the Morality System: Frictionless Purity

One might well think that the quest for ultimate justice that has emerged as the organising force behind the morality system expresses an attractive and even moving ideal, as Williams himself admits (2011, 217). Moreover, the system appears rather well-suited to its task. What, then, is wrong with it? Williams finds numerous things wrong with it, and although he does not always criticize it under that heading, it is clear that much of the criticism levelled either at moral philosophy and ethical theory in general or at Kantianism and utilitarianism in particular applies also to the morality system. One line of criticism is that the morality system leaves us with *too few ethical thoughts and feelings* to be true to ethical experience: like a colour filter laid over the ethical landscape, it masks all but a few morally relevant features of it.[33] Another (and by now rather well-trodden) line of criticism is that the morality system attacks our *integrity* by alienating us from our projects and thus from what sustains the possibility of a meaningful life—the system leaves no-one in particular for

---

[32] Williams suggests that the principle of "ought implies can" derives directly from the fact that moral obligations are understood as *practical* conclusions about what one must do: "If my deliberation issues in something I cannot do, then I must deliberate again" (2011, 195). Of course, as he goes on to note, this does not yet settle the contested issue of what is to *count* as "something I cannot do."

[33] See Williams (2011, 130). For a helpful discussion of this line of criticism, see Krishna (Krishna 2014).

me to be.[34]

But as forceful as these critiques are, I think that the ultimate problem with the morality system, for Williams, is its *frictionless purity*: it robs valuable concepts of their grip on the kind of world we live in, and, by insisting that true value and free agency be pure of any contamination by contingency, threatens to issue in nihilism about value and scepticism about agency. This critique itself has two strands, the *No-Friction* critique and the *Purist-Attitude* critique.

To illustrate the *No-Friction* critique, let us focus on blame$_{(MS)}$. The criticism then is that blame$_{(MS)}$ fails to help us to live in the respects we considered earlier: it cannot help us recruit people into a shared ethical sensibility or bring deviators back into it. The reason is that it fails to get enough of a grip on the kind of world we actually live in, where an agent's character and biographical and historical circumstances largely lie outside that agent's control. A conception of blame that tracks purely voluntary acts has a *null-extension*, so that nothing falls under it. In this sense, it has no friction with, or no empirical basis in, the world we live in. This is a line of criticism we find already in Williams's "The Idea of Equality" (first published in 1962), where he insists that "the concept of 'moral agent', and the concepts allied to it such as that of responsibility, *do and must have an empirical basis*" (1973c, 235–236, emphasis mine). To be concepts worth having, our concepts must allow us to make discriminations *within* the empirical world we live in rather than only between that world and something *beyond* it. This is not the case with blame$_{(MS)}$ and its attendant conception of voluntary$_{(MS)}$ action. To be voluntary$_{(MS)}$, actions would need to be free of, and not to reflect, anything that agents involuntarily and contingently are; but this means that actions can only be voluntary$_{(MS)}$ insofar as agents *chose* the natural endowments, socialisation, education, and other circumstances that shape their lives and their actions. Needless to say, hardly any action will pass this test. We did not choose our circumstances, or if we did, that choice was likely itself a reflection of prior circumstances we did not choose. Voluntariness cannot extend all the way back. Indeed, we *could* not have chosen our circumstances all the way through life, because at the

---

[34] See Williams (1973a, 116–117, 2011, 78, 224). For a recent discussion of the idea that projects are what sustains a meaningful life (though more from a Wigginsian than a Williamsian perspective), see Millgram (2019).

beginning of this process there would have to be the pure, spirit-like self envisaged by the morality system, and this characterless self would lack the basis to make such a choice: it would be too unencumbered by commitments and attachments to get an adequate view of the value of anything.[35]

If, as a matter of fact, the "machinery of everyday blame" (2011, 214) does any work for us at the merely phenomenal level of experience, this is the case because— and only insofar as—it "attempts less than morality would like it to do" (2011, 215). We operate, much of the time, by the lights of blame$_{(ORD)}$ rather than blame$_{(MS)}$. For Williams, that is true more broadly: "If our modern ethical understanding does involve illusions, it keeps going at all only because it is supported by models of human behaviour that are more realistic than it acknowledges" (1993, 11). Blame$_{(ORD)}$ works with a conception of voluntariness that is less demanding than that of the morality system. In inquiring whether an action was voluntary in this undemanding sense, we typically seek to determine only whether the person really *acted*, *knew* what they were doing given the state they were in, and *intended* such and such aspect of what they did (2011, 215–216). This standard—roughly, the standard vindicated already by an explanation of blame$_{(GEN)}$—is one that many actions will easily meet. But if we really consistently allocated blame according to the demanding conception of voluntariness$_{(MS)}$, our blaming practices would largely cease to serve our need for ethical alignment—they would cease to discharge a function that we need to see discharged.

The *No-Friction* critique of the morality system is thus that the purity of blame$_{(MS)}$ and its concomitant ideas robs them of their much-needed friction with the empirical world: too purified to achieve a grip on the rough ground we live on, they become pointless.

It is tempting to conclude that the remedy is simple: we only need to learn to accept the plain fact that the requirements of utter voluntariness cannot be met. We need more truthfulness and knowledge about the world we inhabit, so that we come to see that we are "building ethical life around an illusion" (2011, 212).

But it would be a mistake to conclude that things are so simple. There is a deeper and more insidious problem here, which brings us to the *Purist-Attitude* critique. The problem with the morality system is not just that we fail to find actions that

---

[35] See Williams (1993, 158–9, 2011, chap. 6).

plausibly fall under its purified conceptions; the problem is also that these conceptions shape our attitude towards what we *do* find. The morality system cultivates in its adherents certain *overblown normative expectations* about what shape the world can properly have if value and agency are to have a place in it—the attitude that Williams labels its "purity" (2011, 216). This purist attitude "abstract[s] the moral consciousness from other kinds of emotional reaction or social influence" (2011, 216) and conceives of moral value as lying beyond "beyond any empirical determination" (2011, 217). Its purism lies in its insistence on thinking in terms of stark contrasts between the purity of moral values and free agency and the natural, emotional, and social forces pervading human life when "[i]n truth, almost all worthwhile human life lies between the extremes that morality puts before us" (2011, 216). The demands that the morality system's conceptions make on moral motivation and voluntariness cannot in fact be met. "This fact," Williams writes,

> is known to almost everyone, and it is hard to see a long future for a system committed to denying it. But so long as morality itself remains, there is danger in admitting the fact, since the system itself leaves us, as the only contrast to rational blame, forms of persuasion it refuses to distinguish in spirit from force and constraint. (2011, 216)

This is where the *Purist-Attitude* critique proves crucial: there is danger in admitting to what extent contingency pervades human life as long as one remains attached to the morality system's outlook, because that outlook *blinds* people to the forms of value and freedom that really are to be found in the world we live in.[36]

As a result, disenchanting our view of the world through more truthful naturalistic enquiry risks making things worse rather than better. It risks exacerbating the sense that there is no room for moral value in a world thus understood, resulting in the kind of *nihilism* that maintains that nothing has value. Moreover, because the ethical ideas of the morality system entrained an ethicized psychology reflected in our conceptions of free and rational agency as something that sharply contrasts with and excludes the influence of mere desires and emotions, nihilism about value will

---

[36] As Williams puts it, the morality system "conceals" all the "options for ethical thought and experience" that there are outside itself—"Kantian associations constantly work to short-circuit our understanding" (1993, 77) of those other ways of making sense of things as valuable. This is why, from the perspective of morality, the "Greeks do emerge as premoral" (1993, 77).

be accompanied by *scepticism* about agency. This is why the morality system encounters the problems of free will and determinism "in a particularly acute form" (2011, 195). The system "makes people think that, without its very special obligation, there is only inclination; without its utter voluntariness, there is only force; without its ultimately pure justice, there is no justice" (2011, 218). The result is a bleak and flattening vision of human life, one that elides all difference between rational persuasion and manipulation, convincing and coercing, the force of the better reason and the force of a punch in the face. This danger, which was a central concern of Nietzsche's,[37] and to which Williams gave pride of place in the resounding final lines of the last chapter of *Ethics and the Limits of Philosophy,* shows that merely facing up to the world we live in—merely revising our beliefs—is not enough.

The *Purist-Attitude* critique is therefore this: that in the face of a suitably disenchanted and naturalised view of the world, the attitudes cultivated in us by the morality system threaten to turn us into sceptics or nihilists who see no room in the world either for real value or for real agency. Take, once again, the example of blame. Once one admits to oneself that the demand for utter voluntariness cannot in fact be met, there are two ways one can go: what, in broadly Strawsonian terminology, we might call the *sophisticated optimist* would indeed renounce the demanding conception of voluntariness in favour a less demanding one, voluntariness$_{(MS)}$ in favor of voluntariness$_{(ORD)}$, and confidently allocate blame$_{(ORD)}$ on that basis; this, I take it, is the exit from the morality system that Williams recommends. But it is at least equally tempting to reason in a different direction and to conclude, with the *sophisticated pessimist*, that if the demands of the morality system cannot be met, no act of blaming is ever truly justified, and all we are left with is people being coerced by their circumstances.

Notice that the difference between the optimist and the pessimist is not a difference in knowledge. They both agree that no act is ever truly voluntary$_{(MS)}$ in the way that blame$_{(MS)}$ requires. It is just that while the optimist takes this to speak against that conception of blame, the pessimist takes it to speak against the hope that blame might ever be justified. The pessimist thus betrays his or her continued adherence to the morality system in concluding that since no act is ever voluntary$_{(MS)}$, no ac-

---

[37] See Clark (2015) and Queloz and Cueni (2019).

ceptable form of blame is ever justified—in much the same way that an atheist betrays a residual religiosity if he or she believes, with Ivan Karamazov, that since God does not exist, everything is permitted. The pessimist's position thus involves a kind of *counterfactual adherence* to the morality system (which is structurally analogous to the "counterfactual scientism" (2006c, 187) that Williams accuses Putnam and Rorty of): the pessimist believes that blame is in fact never justified, but that if it *were*, this would *have* to be due to there being utterly voluntary acts.

The pivotal question that separates the pessimist from the optimist may be put as follows: what does the fact that luck and contingency pervade human life *entail*? In drawing from it the conclusion that nothing has value and no-one is free, one betrays a residual commitment to the morality system. In particular, one betrays a commitment to the purist pattern of reasoning encoded in the conceptions of the morality system—a pattern that notably licenses inferences such as the following:

> If anything has value, it is the moral value of things done from moral motives.
>
> If an action is done from a moral motive, it is a voluntary action.
>
> If an action is voluntary, it is not conditioned by anything that is contingent or lies beyond the agent's control.

Via the contrapositives of those claims, one quickly gets from the realisation that every action is in one way or another conditioned by things that are contingent or lie beyond the agent's control to the conclusion that no action is ever voluntary and nothing has value. But as the rather different pattern of reasoning exemplified by the optimist shows, one might also take the same realisation to entail nothing of the sort. Drawing on conceptions of voluntariness, moral motivation, and value that are more tolerant of contingency and draw contrasts such as that between the voluntary and the involuntary, or between the moral and the non-moral, *within* the empirical world we actually find, one can also endorse a different pattern of reasoning—one that allows us to accept that no moral motivation is ever fully pure of contingent desires, or no action ever fully pure of the influence of unchosen circumstance, and still recognise value and freedom in the world. To endorse the first of these two patterns of reasoning rather than the second is not to fall prey to cognitive error; it is to evince a *bad attitude*, an attitude whose badness is *ethical* rather than cognitive. It is an attitude that does not help us to live. As Nietzsche would have put it, it is a life-denying attitude.

We can see that the journey out of the morality system involves not just a first, epistemic step of facing up to reality, but also a second, more radical step, of liberating oneself from overblown normative expectations about just how pure of contingency the world would have to be in order to contain things of value and responsible agents. The conclusion that we are left with *nothing* turns out to depend on an overblown conception of what counts as *something*.[38]

In light of this, it may be helpful to think of the emancipation from the morality system in terms of a movement through three possible stages.[39] At the *initial stage*, one is still blissfully immersed in the illusion that the "rationalistic metaphysics of morality" (1993, 159) correspond to something in reality. Through truthful reflection and inquiry, one moves to a *transitional stage*, where one realizes that those metaphysics do not correspond to anything in reality, but retains the idea that they would *have* to correspond to something if the world *were* to contain true value and free agency, and thus falls into a kind of nihilism and scepticism. Upon being shown that there are other ways of making sense of values and agency in naturalistic but nonetheless vindicatory terms, one can finally move out of the morality system altogether. One "resituate[s] the original opposition[s] in a new space, so that the real differences can emerge" between contingent desires that are moral and contingent desires that are not, between conditioned actions that are voluntary and conditioned actions that are not, and "between the force which is argument and the force which is not—differences such as that between listening and being hit, a contrast that may vanish in the seminar but which reappears sharply when you are hit" (2002, 9). At this third stage, one is liberated from the morality system's constraining conceptions and capable of affirming one's values on different grounds than before. There is no guarantee that *all* of our ideas will survive a truthful understanding of them, but the threat of nihilism and scepticism will have been averted if some of them do. (Arguably, *Ethics and the Limits of Philosophy*, with its withering critique of the various "philosophical errors" (2011, 218) involved in the morality system, primarily helps move its readers from the initial to the transitional stage, whereas *Shame and Necessity* and

---

[38] Williams diagnoses an analogous problem in the view of some ethical theorists that if give up on ethical theory, we are left with nothing (2011, 223).

[39] A comparable schema seems to me to shed light on Nietzsche's conception of the process by which European morality collapses in the wake of the "Death of God." See Queloz and Cueni (2019).

Truth and Truthfulness, with their naturalistic but vindicatory explanations of ideas of agency, responsibility, and intrinsic values like truthfulness, are more concerned to move their readers out of the transitional stage by giving them somewhere outside the morality system to stand.)

The morality system thus turns out to merit its name: it systematically harnesses and adapts to its own ends a variety of initially helpful ideas to hold out the ultimately illusory promise of a shelter from luck. Combining a vindicatory understanding of why we have these ideas in the first place with an initially vindicatory but ultimately critical understanding of why they take the form they do in the morality system, I suggested, can provide us with a nuanced sense of what we need them to do for us, and what kind of friction with the world they need in order to do that. But it is not enough simply to admit that no action is ever voluntary in the sense demanded by the morality system because contingency and luck pervade human life. We also need alternative ways of making sense of value and agency. Only then can we really throw off the powerful misconceptions which, by shaping our sense of how much contingency our ideas of value and agency can tolerate, determine what the pervasiveness of contingency entails.

Bibliography

Candlish, Stewart, and Nic Damnjanovic. 2013. 'Reasons, Actions, and the Will: The Fall and Rise of Causalism'. In *The Oxford Handbook of the History of Analytic Philosophy*, edited by Michael Beaney, 689–708. Oxford: OUP.

Clark, Maudemarie. 2015. 'On the Rejection of Morality: Bernard Williams's Debt to Nietzsche'. In *Nietzsche on Ethics and Politics*, 41–61. Oxford: OUP.

Cueni, Damian, and Matthieu Queloz. Forthcoming. 'Whence the Demand for Ethical Theory?' *American Philosophical Quarterly*.

Fricker, Miranda. 2016. 'What's the Point of Blame? A Paradigm Based Explanation'. *Noûs* 50 (1): 165–183.

Glock, Hans-Johann. 1996. *A Wittgenstein Dictionary*. Oxford: Blackwell.

Greaves, Hilary. 2016. 'Cluelessness'. *Proceedings of the Aristotelian Society* 116 (3): 311–39.

Hartman, Robert J. 2016. 'Against Luck-Free Moral Responsibility'. *Philosophical Studies* 173 (10): 2845–2865.

———. 2017. *In Defense of Moral Luck: Why Luck Often Affects Praiseworthiness and Blameworthiness*. London: Routledge.

Hume, David. 1998. *An Enquiry Concerning the Principles of Morals*. Edited by Tom L. Beauchamp. Oxford: Clarendon Press.

———. 2000. *A Treatise of Human Nature*. Edited by David Fate Norton and Mary J. Norton. Oxford: OUP.

Hyman, John. 2011. 'Action and the Will'. In *The Oxford Handbook of Wittgenstein*, edited by Oskari Kuusela and Marie McGinn, 451–471. Oxford: OUP.

Kitcher, Philip. 2011. *The Ethical Project*. Cambridge, MA: HUP.

Krishna, Nakul. 2014. 'The Morality of Common Sense: Problems from Sidgwick'. DPhil Thesis, University of Oxford.

Latus, Andrew. 2000. 'Moral and Epistemic Luck'. *Journal of Philosophical Research* 25: 149–172.

Levy, Neil. 2011. *Hard Luck: How Luck Undermines Free Will and Moral Responsibility*. Oxford: OUP.

Magee, Bryan, and Bernard Williams. 1971. 'Conversation with Bernard Williams – Philosophy and Morals'. In *Modern British Philosophy*, 150–165. New York: St. Martin's Press.

Millgram, Elijah. 2019. *John Stuart Mill and the Meaning of Life*. New York: OUP.

Monton, Bradley. 2019. 'How to Avoid Maximizing Expected Utility'. *Philosophers' Imprint* 19 (18): 1–25.

Nietzsche, Friedrich. 1998. *On the Genealogy of Morality*. Translated by Maudemarie Clark and Alan J. Swensen. Indianapolis and Cambridge: Hackett.

———. 2001. *The Gay Science*. Edited by Bernard Williams. Translated by Josefine Nauckhoff and Adrian Del Caro. Cambridge: CUP.

———. 2005. 'Nietzsche Contra Wagner'. In *The Anti-Christ, Ecce Homo, Twilight of the Idols, and Other Writings*, edited by Aaron Ridley and Judith Norman, 263–282. Cambridge: CUP.

Otsuka, Michael. 2009. 'Moral Luck: Optional, Not Brute'. *Philosophical Perspectives* 23: 373–388.

Queloz, Matthieu. Manuscript. *The Practical Origins of Ideas: Genealogy as Conceptual Reverse-Engineering*.

———. 2017. 'Two Orders of Things: Wittgenstein on Reasons and Causes'. *Philosophy* 92 (3): 369–397. https://doi.org/10.1017/S0031819117000055.

———. 2018. 'Williams's Pragmatic Genealogy and Self-Effacing Functionality'. *Philosophers' Imprint* 18 (17): 1–20. https://doi.org/2027/spo.3521354.0018.017.

Queloz, Matthieu, and Damian Cueni. 2019. 'Nietzsche as a Critic of Genealogical Debunking: Making Room for Naturalism Without Subversion'. *The Monist* 102 (3): 277–297. https://doi.org/10.1093/monist/onz010.

Ross, W. D. 1930. *The Right and the Good*. Oxford: Clarendon Press.

Sher, George. 2005. 'Kantian Fairness'. *Philosophical Issues* 15 (1): 179–192.

———. 2009. *Who Knew? Responsibility Without Awareness*. New York: OUP.

Statman, Daniel. 2005. 'Doors, Keys, and Moral Luck: A Reply to Domsky'. *The Journal of Philosophy* 102 (8): 422–436.

Walker, Margaret Urban. 1991. 'Moral Luck and the Virtues of Impure Agency'. *Metaphilosophy* 22 (1–2): 14–27.

Williams, Bernard. 1963. 'Freedom and the Will'. In *Freedom and the Will*, edited by David Pears, 1–13. New York: Palgrave Macmillan.

———. 1973a. 'A Critique of Utilitarianism'. In *Utilitarianism: For and Against*, by John Jamieson Carswell Smart and Bernard Williams, 77–155. Cambridge: CUP.

———. 1973b. 'Egoism and Altruism'. In *Problems of the Self*, 250–265. Cambridge: CUP.

———. 1973c. 'The Idea of Equality'. In *Problems of the Self*, 230–249. Cambridge: CUP.

———. 1981. 'Moral Luck'. In *Moral Luck*, 20–39. Cambridge: CUP.

———. 1993. *Shame and Necessity*. Berkeley: University of California Press.

———. 1995a. 'Ethics'. In *Philosophy 1: A Guide Through the Subject*, edited by A. C. Grayling, 545–582. New York: OUP.

———. 1995b. 'Formal and Substantial Individualism'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 123–34. Cambridge: CUP.

———. 1995c. 'How Free Does the Will Need to Be?' In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 3–21. Cambridge: CUP.

———. 1995d. 'Internal Reasons and the Obscurity of Blame'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 35–45. Cambridge: CUP.

———. 1995e. 'Moral Luck: A Postscript'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 241–247. Cambridge: CUP.

———. 1995f. 'Nietzsche's Minimalist Moral Psychology'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 65–78. Cambridge: CUP.

———. 1995g. 'Replies'. In *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams*, edited by J. E. J. Altham and Ross Harrison, 185–224. Cambridge: CUP.

———. 1995h. 'The Point of View of the Universe: Sidgwick and the Ambitions of Ethics'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 153–71. Cambridge: CUP.

———. 1995i. 'Voluntary Acts and Responsible Agents'. In *Making Sense of Humanity and Other Philosophical Papers, 1982–1993*, 22–34. Cambridge: CUP.

———. 1997a. 'Shame, Guilt, and the Structure of Punishment'. In *Scham Und Schuld: Festschrift Aus Anlass Der Verleihung Der Dr. Margrit Egnér-Preise 1997 Zum Thema Scham Und Schuld*, 21–30. Zürich: Margrit Egnér-Stiftung.

———. 1997b. 'What Has Philosophy to Learn from Tort Law?' In *The Philosophical Foundations of Tort Law*, edited by David G. Owen, 487–498. Oxford: OUP.

———. 1999a. 'Bernard Williams'. In *Key Philosophers in Conversation: The Cogito Interviews*, edited by Andrew Pyle, 142–163. London: Routledge.

———. 1999b. 'Seminar with Bernard Williams'. *Ethical Perspectives* 6 (3–4): 243–265.

———. 2001. *Morality: An Introduction to Ethics*. Canto. Cambridge: CUP.

———. 2002. *Truth and Truthfulness: An Essay in Genealogy*. Princeton: PUP.

———. 2006a. 'Moral Responsibility and Political Freedom'. In *Philosophy as a Humanistic Discipline*, edited by A. W. Moore, 119–125. Princeton: PUP.

———. 2006b. 'Pagan Justice and Christian Love'. In *The Sense of the Past: Essays in the History of Philosophy*, edited by Myles Burnyeat, 71–82. Princeton: PUP.

———. 2006c. 'Philosophy as a Humanistic Discipline'. In *Philosophy as a Humanistic Discipline*, edited by Adrian W. Moore, 180–199. Princeton: PUP.

———. 2006d. 'The Legacy of Greek Philosophy'. In *The Sense of the Past: Essays in the History of Philosophy*, edited by Myles Burnyeat, 3–48. Princeton: PUP.

———. 2006e. 'The Primacy of Dispositions'. In *Philosophy as a Humanistic Discipline*, edited by A. W. Moore, 67–75. Princeton: PUP.

———. 2006f. 'The Women of Trachis: Fictions, Pessimism, Ethics'. In *The Sense of the Past: Essays in the History of Philosophy*, edited by Myles Burnyeat, 49–59. Princeton: PUP.

———. 2009. 'A Mistrustful Animal'. In *Conversations on Ethics*, edited by Alex Voorhoeve, 195–214. Oxford: OUP.

———. 2011. *Ethics and the Limits of Philosophy*. London: Routledge.

———. 2014. 'After Virtue: A Study in Moral Theory, by Alasdair MacIntyre'. In *Essays and Reviews 1959–2002*, edited by Michael Woods, 84–86. Princeton: PUP.