https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

MORAL OBLIGATION, BLAME, AND SELF-GOVERNANCE

By John Skorupski

I. Introduction

Moral judgment—the moral assessment of actions and people—is woven inextricably into personal and social life. "Beyond morality" is empty fantasy. A picture of our personal and social world that fails to exhibit this is misleading. How then to make an adequate picture? The moral emotions, their place in social relations, and their potential for both function and dysfunction will have to feature largely. Notions like self-determination, integrity, a common moral sense, and reconciliation with oneself and others will appear. Drawing the picture calls more on the skills of eloquence and hermeneutics than of conceptual analysis. Nonetheless, the latter has a structuring role to play; at any rate, it will be the main concern of this essay.

Two deep features of moral judgment will guide my discussion: the first is that moral judgment is rooted in the rationality of the feelings; the second is that moral agency is self-governing agency.

There are reasons for feelings as well as for beliefs and actions. There can be good or bad reasons for admiration, desire, gratitude, pride; for contempt, anger, resentment, fear. Furthermore, such reasons, which determine the rationality of feelings, are not reducible to epistemic reasons (reasons to believe something) or practical reasons (reasons to do something). I shall call them *evaluative* reasons. Moral propositions are propositions about evaluative reasons: specifically, about reasons for the sentiment of blame. Understanding the moral thus requires that we describe this sentiment without circularity and in sufficient depth.

This can be done because blame, like other sentiments, has characteristic objects and characteristic actions to which it disposes. We can identify the sentiment by describing them. In this essay, however, my focus will be almost entirely on object rather than action. I shall only sketchily describe the action to which moral blame disposes—*withdrawal of recognition.* The significance of this action is a subject that requires full treatment in its own right.[1]

[1] Some further discussion can be found in John Skorupski, *Ethical Explorations* (Oxford: Oxford University Press, 1999), part III; and Skorupski, "Blame, Respect, and Recognition: A Reply to Theo van Willigenburg," *Utilitas* 17 (2005): 333–47.

doi:10.1017/S0265052509990197

© 2010 Social Philosophy & Policy Foundation. Printed in the USA.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

My first aim is to show how moral wrongness is definable in terms of evaluative reasons to blame, and how these reasons connect to practical reasons. This will require us to take account of the second deep feature of the moral: that moral agency is, by its very nature, a species of self-governing agency.

Self-governing actors can act from reasons. To do so, they must be able to think about what reasons they have. They must be able to audit those reasons reflectively, in order to assess whether they have sufficient reason to believe, or feel, or act—or whether they must investigate further before they have sufficient reason. And they must be able to act on their conclusions.

Beyond these indisputable points, however, we shall see that self-governance can be understood in weaker and stronger ways. Moral agents require at least the capacity for *self-determination.* This is a two-fold capacity: (i) the power to think about their reasons and determine whether they are sufficient or whether further inquiry is needed; and (ii) the power to act on their conclusions. I am self-determining if I determine myself by propositions about reasons which I myself accept—whatever the source of my acceptance. But self-governance is often understood to include not just self-determination but also *autonomy.* The idea of autonomy, understood in a broadly Kantian way, is stronger than that of self-determination; it is self-determination grounded in personal normative insight into truths about reasons: the ability to recognize such truths by one's own lights, in a first-person way. My second aim is to consider whether moral agency presupposes the weaker notion of self-determination or the stronger notion of autonomy. Modern ideals of morality place great weight on the idea that every moral agent is equally autonomous; but we shall see that moral agency requires only the capacity of self-determination, and that the idea that all moral agents are equally autonomous is hard to sustain.

In Section II, I define some concepts we shall need and summarize some structural theses about reasons. Sections III and IV examine the connections between moral wrongness, blame, and reasons. Sections V and VI discuss self-determination and autonomy. Section VII summarizes my conclusions.

II. Reasons and Self-Governance

We start, then, from the fact that there are three kinds of reasons: evaluative, epistemic, and practical. They have a largely common logical structure; in all cases, reasons are relations between facts, persons, and "responses" (feelings, beliefs, actions). Furthermore, in all these cases, three distinct reason-relations fall under the concept of a reason:

(i) One can say that some particular facts are a reason, weaker or stronger, for some person at some time to ψ (where ψ ranges

over beliefs, feelings, and actions). I call such facts *specific* reasons (R): facts π_i are at time t a reason of degree of strength d for x to ψ—$R(\pi_i, t, d, x, \psi)$.

(ii) One can say that taking all the facts that count for and against into account, there is more or less strong reason *overall* for some person at some time to ψ. Such facts constitute an *overall* reason (R_o): facts π_i are at time t overall reason of degree of strength d for x to ψ—$R_o(\pi_i, t, d, x, \psi)$.

(iii) One can say that some facts give some person at some time *sufficient* reason to ψ. So we also have sufficient reasons (S): facts π_i are at time t a sufficient reason for x to ψ—$S(\pi_i, t, x, \psi)$.

When there is sufficient reason for x to ψ, we can say that x should* ψ. (I give this "should" an asterisk because while this is certainly one way we use "should," it has plenty of other uses.)

However, there is also an important difference between epistemic reasons and practical and evaluative ones. The facts that give you reason to believe must be facts that you know or could know, whereas that restriction does not apply to facts that give you reason to feel something or do something. If the building is about to explode, that is a reason for you to leave, whether or not you know it or could know it. Likewise, if a benefactor is anonymously helping you, that is a reason for you to be grateful to that person, even though you cannot find out who it is. But now suppose that a hurricane will strike London next year. Is that a reason for you to believe that a hurricane will strike London next year? No. Reasons to believe that proposition, now, must consist of evidence, or more generally, of facts that you could now come to know, whether or not you actually know them. Think of these facts as your *epistemic field;* your epistemic reasons (R, R_0, S) are relative to your epistemic field. Spelled out in full, therefore, these reasons have your epistemic field as one of their relata, whereas your evaluative and practical reasons do not.

We come now to a distinction that is critical for our purposes. Self-governance requires that one should be able to tell what reasons one has. You have to be able to reflect on your specific reasons, estimate how strong they are overall, and determine whether they are sufficient. Thus, in an example given by Allan Gibbard,[2] the way out of the forest that you *should** take is the one that is in fact the quickest: that is what there is sufficient reason for you to do. But you are lost; you don't know the quickest way, and there is no one to tell you. So what you actually have reason to do is, say, pick an arbitrary straight line and follow it.

As this example illustrates, we talk about reasons in two ways (among others), depending on whether or not we are thinking of them as *self-*

[2] Allan Gibbard, *Wise Choices, Apt Feelings* (Oxford: Oxford University Press, 1990), 18–19.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

accessible. A way to mark this contrast, as I have just done, is to distinguish between

(i) the reasons there *are* for a person to ψ

and

(ii) the reasons a person *has* to ψ.

On this way of speaking, the reasons there *are* for a person are not in general self-accessible. He may need to investigate his epistemic field to discover them, or they may lie outside his epistemic field. In contrast, the reasons a person *has* are self-accessible: if he has them, he can know—recognize, tell—that he has them, solely by means of reflection and self-examination, including attention to the way things seem to him. (The levels of normative insight, cognitive competence, and self-scrutiny needed may require time and capacities that he does not have; the point is that nothing *else* is required.)

However, we also talk about "the reasons a person has" in other ways. Quite often, for example, we are not referring to a person's warranted reasons but to the reasons he *takes* himself to have. Or we may be referring to the reasons there would be for him if his factual beliefs were true. To avoid ambiguity, then, I shall talk about a person's self-accessible reasons as *warranted* reasons. And I shall call a warranted *sufficient* reason a warrant. When I refer to a person's reasons simpliciter, I am referring to the reasons there are for that person, not to the warranted reasons he has.

When we talk about the reasons a person has to ψ at a given time—in the sense of his warranted reasons—we are referring to what there is sufficient reason for him to believe are the reasons for him to ψ, on the sole basis of facts knowable by self-scrutiny at that time. (He may or may not actually form those beliefs, or come to know those facts.) These facts are a subset of his epistemic field: call them his *epistemic state.* Warranted reasons are then definable in terms of reasons: a person's warranted reasons to ψ are the reasons to ψ that there is sufficient reason for him to believe obtain, relative to his epistemic state. Thus, there is warrant for x to ψ at t if and only if, relative to x's epistemic *state, x* should* believe that x should* ψ at t.

Now if I can have knowledge of my reasons, it must be possible for me to tell, just by normative reflection—pure reflection about reason relations—that some facts would constitute a reason to ψ. Suppose one can tell solely by such reflection that were π_i to obtain, it would be the case that $R(\pi_i, x, \psi)$. Then we can say that if these facts obtained, they would be a *complete* specific reason for x to ψ. (Likewise for overall and sufficient reasons.)

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

In putting it this way, I am not assuming a particular meta-normative view about the epistemology and ontology of reasons. One can be a cognitivist or a noncognitivist with respect to claims about complete reasons (i.e., hold that they assert true or false propositions, or hold that they express affective attitudes) while accepting that agency requires that agents must be able to recognize, accept, or endorse complete reasons by pure reflection rather than empirical enquiry, if they are able to assess what warranted reasons they have at all.

These complete reasons may be very detailed or particularistic. For example, the fact that the house is about to explode may be a reason for you to leave it or a reason for you to enter it. It depends on whether you are a bomb-disposal expert. And if you are, then it also depends on other things, such as whether you are there in your official capacity, whether you are non-exploitatively employed, whether it is possible for a trained professional to go in without excessive risk, whether you are sufficiently trained, and so on. Any of these facts could be cited as a reason in an appropriate conversational context. But there must be a total set of facts in virtue of which the cited fact, whatever it is, can be said to be a reason. In general, there must be sets of facts that constitute complete reasons if there are reasons at all.

Self-governance, even as self-determination, presupposes the ability to recognize complete reasons (though not necessarily to state them). A Kantian argument from *autonomy* goes even further. It presupposes that complete reasons take the form of pure normative truths, and that self-governing agents have *a priori,* first-person insight into these truths. We shall come back to this.

III. Wrongness and Blameworthiness

But let us now turn to moral obligation. Saying that *x* has a moral obligation to α is clearly not identical in meaning to saying either that *x* should* α or that *x* has a warrant to α. It can be true that you should* α, true that you have a warrant to α, but false that you have a moral obligation to α. The converse is less plain: can it be true that you have a moral obligation to α but false that you should* α or that you have a warrant to α? We shall answer that question.

To say that you have a moral obligation to α is to say that it would be morally wrong for you not to α. There is then a connection between moral wrongness and blameworthiness. However, this connection is not straightforward. I take it to be as follows:

(1) It is morally wrong for *x* to α if and only if, were *x* to α from the beliefs that are warranted in *x*'s epistemic state, then either *x* would be blameworthy for α-ing or extenuating circumstances would apply to *x*'s α-ing.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

Obviously, this requires explanation. Extenuating circumstances remove or diminish blame that would otherwise have been appropriate. The reference to action from the actor's actually warranted beliefs arises from the self-governing character of moral agency and will be explained and defended in the next section. But first let us consider what 'blameworthy' means.

There is a causal meaning of 'blame' in which one can blame the dud calculator-batteries for a mistake in the bill, blame the weather for the train delay, etc. Saying that they are to blame is just saying that they are the proximate cause of the unfortunate result. It is not this causal sense of 'blame' that is in play in thesis (1). There is also a moral sense. One can say, for example, both that Mary was to blame for the misunderstanding, and that she should not actually be *blamed*—she was justifiably attending to something else that was happening. She was to blame in the causal sense, but not in the moral sense. Blame in (1) is to be understood in the moral sense.

Now when we talk of blaming someone in this sense, we are normally referring to a judgment, or an action that is the expression of that judgment. The judgment is a judgment about the appropriateness of a sentiment—the feeling or sentiment of blame. To judge that someone is morally blameworthy is to judge that the blame-feeling toward them is reason-supported, that there is sufficient reason for it, whether or not one actually feels it. The act of blaming is the act of expressing that judgment. For convenience, however, I am going to use the term 'blame' to refer directly to the feeling, not to the judgment or the act. Thus, in (1), 'blameworthy' means 'blame-feeling-worthy': there is sufficient reason for the sentiment of blame.

It may happen, for a variety of reasons, that though the blame-feeling is reason-supported and the judgment that it is is warranted, it should not be expressed. It may be counterproductive to do so, for example, in getting someone to change their ways. It is thus important to keep clear the distinction between the sentiment and the act of blame. It is also important in other, more theoretical ways. When John Stuart Mill gives his well-known 'blameworthiness' definition of moral wrongness in chapter 5 of *Utilitarianism*,[3] it is unfortunately not clear whether he is concerning himself with the sentiment or the act of blame. However, inasmuch as he commits himself, in theory at least, to assessing the appropriateness of blame by utilitarian standards, he must be interpreted as referring to the *act* of blame. For utilitarianism is a view about practical reasons, reasons for actions, not about evaluative reasons, reasons for feeling. Insofar as some utilitarians elide this point, they do so because they do not really believe in evaluative reasons, or they assimilate them to practical reasons. In reality, however, when we assess whether the *sentiment* of

[3] John Stuart Mill, *Collected Works* (London: Routledge, 1963–91), vol. X, 246.

blame is reason-supported in a particular case, we are not considering the expediency of the *act* of blame. Properly moral judgment is, rather, a special case of our general evaluative capacity to assess reasons for feelings as such. In that respect, the view proposed here is more like Adam Smith's than like Mill's, in that Smith takes the "propriety" of feelings in general to be determined by their objects, not by the utility of the actions they dispose agents to take.[4] This divergence makes a very big difference.

Granted these elucidations, I claim that (1) is true *a priori.* Can it in that case serve as a *definition* of 'morally wrong'? There is a "cart and horse" objection to any proposed definition of moral wrongness in terms of the blame-feeling. Thesis (1), it may be granted, is an *a priori* truth; but it is so by definition of 'blame,' not by definition of 'morally wrong.' Blame in the moral sense is defined as that sentiment which there is reason to feel toward something that is morally wrong. So (1) is true by definition, but by definition of 'blame.' 'Morally wrong' is semantically primitive in the sense that it is indefinable.

I think this makes a correct point about, so to speak, the lexicon of English. But I don't think it shows that the concept of the morally wrong cannot be characterized in terms of the concept of an evaluative reason for blame. We can elucidate a concept by saying what is required for possession of the concept, rather than by the semantic route of explicit definition within the existing resources of a language. What you distinctively need, to have the concept of moral wrongness, is sensitivity to reasons for the blame-sentiment.

On this approach, however, we cannot characterize blame as 'that sentiment which there is reason to feel toward something that is morally wrong.' To be fruitful, this approach must characterize the blame-sentiment without appealing to the notion of moral wrongness. How to do this?

In general, we characterize emotions by the objects that arouse them and the actions which they prompt. We may indicate the objects by showing exemplary cases, and where possible, by stating some general condition on the objects. Thus, we could individuate the blame-sentiment by saying that it is what one is disposed to feel toward an agent in a variety of sufficiently well-described cases—for example, toward someone who steals, who acts cruelly, who shows gross ingratitude. Furthermore, a general condition will be important: the object of the blame-sentiment must be something which the agent had warranted reason not to do, and which the agent could have refrained from doing. (I discuss this further in Section IV.) This is, of course, not sufficient: the nature of the reasons matters. It is the fact that the agent did the thing, despite *those* reasons, reasons of kinds which we exemplify by giving examples, that gives sufficient reason for responding with the sentiment of blame.

[4] Adam Smith, *The Theory of Moral Sentiments* (Indianapolis, IN: Liberty Fund, 1984), Part I, section 1, chapter iii, and Part II.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

We can also approach the blame-sentiment from another direction—from the characteristic action which it prompts.

Blame does not prompt an agent to attack (as anger does), or to flee (as fear does). It is not resentment, because resentment is specifically occasioned by what is taken as injury to oneself. Gratitude and resentment are agent-relative feelings, constituting the realm of benefits and torts, whereas the blame-sentiment is agent-neutral and constitutes the realm of right and wrong. If there is reason for anyone to feel it toward the object, there is reason for everyone to feel it. Its characteristic behavioral quality is a chilliness or a shrinking away or withdrawal, rather than aggression. This is particularly clear in one's own case, in the experience of self-blame or guilt—which could hardly be resentment directed against oneself, but is indeed a shrinking from oneself. (The resentment view must make guilt something different from blame: guilt might be viewed as fear of the resentment and retaliation of others, for example.)

Blame disposes agents toward *withdrawal of recognition.*[5] To withdraw, withhold, or refuse recognition is to cut off relations or refuse to enter into them, to exclude, in however partial and temporary a way; in more extreme cases, it leads to ostracism, casting out, outlawing. Like fear, blame disposes one to create distance between oneself and the object; but the difference is that the disposition of fear is to flee, whereas the disposition of blame is to cut off or exclude. Guilt, or self-blame, is the withdrawal of recognition from oneself. (Hence, "I couldn't live with myself if I did that.") And it is important that recognition can be restored through a process of reconciliation that typically involves remorse, repentance, and forgiveness.

This sketch of the blame-sentiment plainly needs to be filled in. But let us assume that a fuller account can fix the sentiment clearly enough to allow the non-question-begging introduction of a term that denotes it: 'β'. The referent of 'β' is the sentiment that (i) is the appropriate, reason-supported response to certain paradigms of belief and action, and (ii) prompts a characteristic type of action, namely, withdrawal of recognition. We then have the following definition:

(1ª) It is morally wrong for x to α = were x to α from the beliefs that are warranted in x's epistemic state, and without any extenuating circumstances, there would be sufficient reason to feel sentiment β toward x for α-ing.

Thesis (1) should now be understood in the sense of (1ª). Furthermore, this definition can be widened into a "sentimentalist" account of moral

[5] Compare T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, MA: Belknap Press of Harvard University Press, 2008), chap. 4.

concepts in general, that is, a definition in terms of evaluative reasons, reasons for feeling.[6]

IV. Morality, Blame, and Practical Reasons

If I blame someone for doing something, I must think he had a reason not to do it. That, though not sufficient, is necessary: if you show that he had no reason not to act as he did, you refute the charge of blameworthiness. There can be no reason to blame someone who *had no reason not to do* what he did.

This is true, but in what sense? As noted in Section II, talk of the reasons a person has is ambiguous. In blaming someone, I am not directly concerned with his warranted reasons; I am concerned with the reasons there would be for him if his factual beliefs were true. Let us call these the agent's belief-relative reasons.[7] Belief-relative reasons are not reasons in either the primitive sense or the warranted sense.

Blame does not follow reasons as such. Suppose you have just inadvertently taken poison. There is reason to ring for an ambulance. But I am not blameworthy for failing to call an ambulance if I have no warranted reason to believe that you have taken poison, and thus fail to believe that you have.

Nor, however, does blame follow warranted reasons. Suppose that I have warrant to believe that you have just swallowed poison, but, failing to recognize that warrant, I fail to believe that you have, and do nothing. In this case, I am not blameworthy unless I could have been expected to recognize that warrant for belief. If my failure to recognize the warrant is negligent, then I may be blameworthy; the blame attaches to the failure to recognize the warrant, not to the failure to help you. Note that this failure is still practical: it is a failure to think carefully enough, for example, about what was in that packet, and that is a failure to do something: "you didn't think." If, however, in the circumstances I could *not* have been expected to recognize the warrant—there was no negligent failure to think—then I am not blameworthy.

In contrast, suppose that I believe, though without warrant, that you have taken poison, and I do nothing. Am I blameworthy? Yes. We could quite naturally say that given that belief, I had reason to call an ambulance, and here the sense of 'had reason' is the belief-relative sense mentioned above: there would have been reason for me to call an ambulance if my beliefs were true.

[6] I argue that this is so in my book *The Domain of Reasons* (Oxford: Oxford University Press, forthcoming). The themes of this essay will be developed more fully there.

[7] Note the restriction to factual beliefs. Purely normative beliefs that are false can sometimes excuse or remove a person from blame; however, in the following examples, I am abstracting from such issues.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

Next, suppose you administer a drug to a patient, and given your epistemic state, there is warrant for you to think it is a cure. But in fact you don't believe that; you *unwarrantedly* believe it to be a poison.[8] You took yourself to be poisoning the patient; in fact, you cured him. Let us further suppose that other facts that you believe to obtain would not, if they did obtain, give you sufficient reason to poison the patient (for example, you do not believe him to be Hitler, etc.). Are you blameworthy? You are, just because you believed yourself to be poisoning the patient. So either that was what you intended to do, or you didn't care that that would be (as you thought) the effect of your action. Either way, you are blameworthy; yet there was no *warranted* reason for you not to administer the drug. Because blame follows the beliefs under which you acted, it follows neither the reasons there were for you, nor the warranted reasons there were for you, but your belief-relative reasons. In this case, there was belief-relative reason for you not to administer the drug.

Thus we have:

(2) If your α-ing is blameworthy, then if your factual beliefs were true, there would be reason for you not to α.

This feature of blame, its connection to your belief-relative reasons, is inherent in the sentiment itself: one grasps it in understanding the sentiment of blame and thus understanding what can constitute an intelligible object of the sentiment. It makes no sense, lacks reason, to blame people for doing what they had no belief-relative reason not to do (though they may be blameworthy for having the beliefs they have). This is a point about the immanent rationality of feelings: it makes no sense to blame in such cases, in the way that it makes no sense, lacks reason, to resent a good turn without in any way thinking it to be maliciously meant, patronizing, etc. Someone who resents a good turn has to read malice or slight into the other's action, and similarly, someone who feels the sentiment of blame toward the other for what he did must read some belief-relative reason for not acting into the other's state of belief, or some epistemic negligence ("He knew, *really*"; "He *should* have realized"; etc.).

So far, we have been considering blameworthiness, not moral wrongness. Blame connects to belief-relative reasons; thesis (1), however, connects moral wrongness to *warranted* beliefs about the facts.

I submit that the actual beliefs from which an action was done are relevant to its blameworthiness but not, in general, to its moral permissibility or wrongness.[9] Say you administer the drug when there is actually

[8] The example is adapted from an unpublished discussion of these issues by Derek Parfit, which I have found very helpful.

[9] If I believe that an action is morally wrong, it does not follow that it is. But, depending on my actual factual and normative beliefs, it may be blameworthy on my part to do it, even

warrant, given your epistemic state, for you to believe that it will cure the patient; but you fail to recognize that warrant and, in fact, give it to him from the *unwarranted* belief that it will poison him, with the intention of poisoning him. In that case, your action is blameworthy (absent extenuation), but according to thesis (1) it is *not* morally wrong. Had your belief been warranted, you would have been doing something morally wrong. In fact, it was not warranted, so although you acted with a blameworthy intention, the action itself was not morally wrong. That remains the case even if, despite the fact that you have warrant to believe the drug will cure the patient, the drug actually kills him. So you intended to kill him, and you killed him: even so—though you are much to blame for what you did—what you did, given your epistemic state, was not morally wrong. Given what you had warrant to believe, it was actually the morally right thing to do.[10]

Moral wrongness turns neither on the agent's actual beliefs, nor on the facts, but on the beliefs warranted in the agent's epistemic state. Although blame, in contrast, is primarily concerned with the actual beliefs under which you acted, there is nonetheless a connection between blameworthiness and warranted reasons, as follows:

(3) If, were *x* to α from the beliefs that are warranted in *x*'s epistemic state, either *x* would be blameworthy for α-ing or extenuating circumstances would apply to *x*'s α-ing, then *x* has warranted reason not to α.

Suppose, first, that *x* would be blameworthy for α-ing. Then, by thesis (2), *x*'s factual beliefs must be such that, if they are true, there is reason for *x* not to α. Now suppose those factual beliefs that *x* has are warranted; then *x* has warranted reason not to α. We get to this conclusion from the self-determining nature of moral agency: if some facts would constitute a reason for a self-determining agent *x* not to α, then *x* can tell by reflection alone that they would. Thus, if he is warranted in believing that those facts obtain, he has warrant to believe that there is reason for him not to α; that is, he has a warranted reason not to α.

Suppose, secondly, that extenuating circumstances apply to *x*'s α-ing. Extenuation exists when an action that would normally be blameworthy is excusable. It may be excusable *either* because you excusably failed to become aware of warranted reasons not to do the action *or* through some excusable failure to act on those reasons, even though you were aware of them (for example, the understandable emotional turmoil you were in).

if it is not morally wrong. Furthermore, conscientious conviction can itself affect moral permissibility.

[10] I am grateful to David Copp for very helpful discussion of this point (about which we disagree).

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

These alternatives cover large categories. The first, in particular, may involve something as simple as the fact that there was a great need for haste and you excusably failed to consider a possible consequence, or something as complex and hard to evaluate as the fact that in the society in which you live, you could not be expected to grasp the reasons not to do what you did. But given that this is the overall shape of extenuation, to say that extenuating circumstances apply to a person's action implies that there *was* warranted reason for that person not to do the action. Thus, we arrive at (3).

From (1) and (3), we conclude:

(4) If it is morally wrong for x to α, then x has warranted reason not to α.

I suggest, further, that (3) can be strengthened. Can there be reason to blame a person who, acting from warranted beliefs about the facts, does what he had warrant to do? Surely not. You cannot be blameworthy for doing what you had sufficient reason to do because you correctly saw that you had sufficient reason to do it: an action done from warrant cannot be blameworthy. This gives us:

(5) If, were x to α from warranted factual beliefs, either x would be blameworthy for α-ing or extenuating circumstances would apply to x's α-ing, then x has no warrant to α.

Thesis (5) seems to me to arise, again, from the nature of the blame-sentiment; like thesis (2), it is a principle internal to the sentiment of blame, a principle governing the intelligibility of that sentiment. How could there be reason to blame you for acting from warrant—doing what you had most reason to do because it *was* what you had most reason to do? What is there for blame to latch onto? You could *reasonably* be blamed only if you could and should have believed that there was more reason for you to do something else, that is, that there was warrant not to do what you did. So, for example, suppose it were true that you always have sufficient reason to do what you warrantedly believe is best for you—as egoism says. Then it could not be *blameworthy* for you to do that.

By a similar argument to the one just given, thesis (5) yields:

(6) If it is morally wrong for x to α, then x has no warrant to α.

Since it is morally obligatory not to α if and only if it is morally wrong to α, we can conclude from (4):

(7) If it is morally obligatory for x not to α, then x has warranted reason not to α.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

Likewise, we can conclude from (6):

(8) If it is morally obligatory for *x* not to α, then *x* has no warrant to α.

Can we further deduce that *x* has warrant not to α? If (acting from warranted factual beliefs) you were either blameworthy for α-ing or extenuation applied, you must have had warranted sufficient reason to believe that there was sufficient reason not to α.[11] That is, you must have had warrant not to α. Thus, as well as (8), we have:

(9) If it is morally obligatory for *x* not to α, then *x* has warrant not to α.

Thesis (8) and thesis (9), taken together, say that moral obligation is *categorical*, in the sense of being inescapable and supreme.[12] That is, if you have a moral obligation to do something (not to omit doing it), then you have warranted sufficient reason to do that and no other thing. This conclusion is striking: not merely is the categoricity of morality *consistent* with a sentimentalist account of moral concepts in terms of reasons to blame—it actually follows from it. The key to the categoricity of morality lies in the hermeneutics of the blame-sentiment.[13]

My account of moral obligation relativizes it to epistemic states. We can confirm the account, in the first place, by examples. There is the example already mentioned: if I have warrant to believe that you have unwittingly taken poison, then I have a moral obligation to ring for the ambulance. That is so, even if in fact you have not taken poison.[14] In contrast, if you have unwittingly taken poison, but I have no warranted reason to believe this, then I have no moral obligation to ring for an ambulance: I do not act morally wrongly if I do not do so.

Consider now the case of the intended poisoning that cures the patient. Here moral wrongness and blameworthiness intelligibly diverge. Suppose the agent had warrant to believe that administering the drug would

[11] That is, not just *insufficient* reason to believe that there was sufficient reason to α (which would also entail absence of warrant to α).

[12] This is David Brink's terminology. See David O. Brink, "Kantian Rationalism: Inescapability, Authority, and Supremacy," in Garrett Cullity and Berys Gaut, eds., *Ethics and Practical Reason* (Oxford: Oxford University Press, 1997).

[13] Moral obligation is categorical, but that is not to say that it is reason-constituting. If you have a moral obligation to α, you have warrant to α; but it is not the fact that you have a moral obligation to α that constitutes that warrant. The moral obligation exists *in virtue of the nature of the warrant*. It is because of the kind of reasons there are for doing the thing that it is morally obligatory to do it. Those reasons are what make it morally obligatory—moral obligation supervenes on them: that is, knowingly to ignore them merits blame.

[14] It is true that, if you do not die, people would not in practice feel the sentiment of blame as strongly as if you had. That is a significant point; however, I think the general view would still be that I had a moral obligation to act and was blameworthy not to have acted.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

cure the patient; in that case, I argued, it was not morally wrong to administer it. That conclusion remains true even if the agent nonetheless *unwarrantedly* believed that the drug would poison the patient and was thus blameworthy for administering it, and even if, in fact, the drug did kill the patient. Symmetrically, if the agent had warrant to believe that the drug would poison the patient, it was morally wrong to administer it, even if the agent unwarrantedly believed that the drug would cure the patient and in fact it did. In the first case, we have blameworthiness without moral wrongdoing; in the second, we have blameless wrongdoing.

It is true that if you are offering me moral advice, you might well say to me, "It would be morally wrong to do that," and then mention some facts—which I could not otherwise have known—that (as we would naturally say) *make* it morally wrong. Thus, it seems that what makes an action morally wrong is the facts, not the agent's warranted beliefs about the facts.

But we need to consider the requirements of advice. When I ask for advice about what I ought to do, I *may* be asking what I am warranted in doing in my actual epistemic state. This is so, for example, if I know that you and I are in the same epistemic state. Or if I am in a group of people who share an epistemic state, then my question "What ought we to do?" is the question "What are we *warranted* in doing?" But if I ask someone who knows more, then I am asking for advice in the light of what *he* knows. Thus, in Allan Gibbard's forest example, if I ask a forester what we ought to do to get out of the forest, I ask because I hope he knows the best way. If he does, then it would be wholly inappropriate for him to tell me to take a straight line and stick to it, even though that is what I am currently warranted in doing in my present state of ignorance.[15] He should be helping me by advising me in the light of what he knows, not what I know. Exactly the same applies if your advice is about the moral aspects of my decision. You should be advising me about those aspects in the light of what *you* know, not what I know.

Speaking on the basis of what he knows, the forester might simply say, "This is the best way out," without telling me why. In telling me that, he would not be implying that I already had a warrant for thinking this is the best way out. But at this point a contrast emerges. What if you said, "This is what you have a moral obligation to do," without telling me why? Strictly speaking, that would be true only if I *already* had a warrant for thinking I had that moral obligation. If you were basing your advice on facts I did not know, you should have made that clear.

Suppose, however, that for one reason or another, confidentiality perhaps, you are unable to tell me the relevant facts. What you could still say is, "If you knew the relevant facts, you would have such and such a moral obligation." Next, suppose I am warranted in thinking that you are reli-

[15] As Gibbard points out (*Wise Choices, Apt Feelings*, 19).

able. In general, agents have a moral obligation to do that which someone they know to be reliable assures them they would have a moral obligation to do if they knew the facts. The obligation arises because, in virtue of what the reliable person has said, they have sufficient reason to believe that the facts are such as to constitute sufficient reason to do the thing—a reason, moreover, that it would be blameworthy to ignore if they knew what it was. It is surely blameworthy not to act on such a warrant.

For these reasons, it is natural to formulate moral principles in terms of the facts: "If a person has taken poison, you ought to call an ambulance," not "If you have sufficient reason to believe that a person has taken poison, you ought to call an ambulance." When we formulate principles in this way, we are spelling out the reasons which, if you know of them, it is blameworthy on your part to ignore. It is in this sense that moral wrongness supervenes on the facts.[16]

V. Morality, Self-Determination, Autonomy

A survey of our judgments about moral obligation and blame seems to me to tell in favor of the foregoing account of moral obligation. But there is a more general consideration. Moral obligations, like rational requirements, follow warranted beliefs, because we must be able to *act from* them. People must be able to determine reflectively what moral obligations they have in a concrete situation, just as they must be able to determine reflectively what specific warrants they have. Similarly, the connection between failure to comply with a moral obligation and moral criticizability is like that between failure to comply with a warrant and rational criticizability. In both cases, we need to know the actor's warrants, actual beliefs, and excuses. We can ask what a person was either *rationally* or *morally* required to do given the factual beliefs he actually held, and we can ask whether he had warrant for his factual beliefs. We can also ask whether he was rationally or morally required, in the circumstances, to pay more attention than he did to establishing warrants. In short, the notions of moral obligation and rational requirement occupy the same role-position in the pattern of normative assessment.

As I said at the outset, however, we should distinguish between self-determination and autonomy proper. Self-determination consists in recognizing what one has sufficient reason to do relative to one's acceptance of certain normative tenets, and acting from that recognition. I am self-

[16] Final elucidation: I have been speaking about moral obligations, not about rights and the corresponding duties of right. The latter should be distinguished from moral obligations. Suppose you are a deceased person's executor, and that person has left a last will. You don't know that: the will is lost but still exists. The will leaves the property to X. In that case, X has a right to the property and you have a duty of right to execute the property in accordance with the will. But you don't have a moral obligation to do so, because you don't know of the will. (You may have a moral obligation to make a search, etc.)

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

determining if I determine myself by normative tenets which I personally accept, and do not merely "know about" (e.g., as being generally accepted, or dangerous to ignore, etc.). I am self-determining even if I accept those tenets themselves on authority, on trust, or by my own *fiat*, and whether or not I have personal insight into their truth (indeed, even if they cannot be said to *be* true or false). I can determine myself, act from my own assessment of reasons, so long as I am able to audit my epistemic state, determine what reasons I have in the light of my acceptance of certain normative propositions or commitments, and act from those reasons. The same distinction applies in the moral case. I am morally self-determining if I am able to audit my epistemic state, determine the concrete moral obligations I have in the light of my acceptance of certain moral tenets, and act from those obligations. I can have this ability even if I have no personal insight into moral truths, and even if there are no moral truths. What is required for self-determining moral agency is simply that I am able to audit, from within my epistemic state, what specific moral obligations I have in a concrete situation—relative to moral tenets I personally accept.

It is inherent to the very notion of moral agency—with its dimensions of answerability and susceptibility to blame—that it is at least self-determining agency. It follows that the notion of self-determination is present wherever moral concepts are found. In contrast, the idea of moral *autonomy* is stronger. Moral autonomy is more than self-determination; it is self-determination grounded in personal moral insight, that is, in the ability to recognize, by one's own lights, true moral propositions as true. The same can be said of rational autonomy; it is self-determination grounded in personal insight into reason relations as such, that is, the ability to recognize personally the truth of true propositions about reasons.

We have weaker or stronger notions of rational and moral agency, depending on whether we think of these forms of agency as merely self-determining or as requiring full autonomy. The stronger notion, that of autonomy, clearly assumes that normative tenets are indeed *propositions,* whose truth or falsity is knowable by personal insight. I will call the principle of moral or rational insight that is assumed in this notion of autonomy, the (moral or rational) *Insight Principle.* For the moral case, it can be stated as follows:

> *Insight Principle:* Moral agents can know their moral obligations (in all specific, concrete situations) by means of their own insight into them.

For rational autonomy, exactly the same can be said, substituting "rational" for "moral," and "warrants" for "moral obligations." Autonomy in both cases can be a matter of degree, depending on how much insight actors have, and how much control over their acts.

Rational autonomy can furthermore be understood broadly or narrowly. The broad understanding takes rational autonomy to include a grasp of evaluative reason relations. The narrow understanding takes it to include only epistemic and practical reason relations. Since rational autonomy consists (in the ideal case) of first-person insight into *all* reason relations, if we accept that there are evaluative reasons we must understand it in the broad way. Moral autonomy is then a special case of rational autonomy; it calls on broad, not narrow, rationality. It is insight into reason relations, but the reason relations involved include evaluative reason relations. So I agree with Kant that moral autonomy is rational agency. But because Kant did not accept that there are evaluative reasons at all, the difference between the picture of morality sketched here and Kant's picture of it is great.

However, this difference, albeit great, is independent of Kant's master thesis, that moral agency requires autonomy. That could be true on either picture of morality. And this thesis, the thesis that human beings are moral agents only because they are autonomous, has fundamentally shaped modern ethical and political ideas. Particularly influential is an *egalitarianism of autonomy,* according to which all human beings are equally autonomous; and this is especially held to apply to *moral* autonomy. Nothing in the ethical legacy of Kant and Rousseau is more influential than this: the idea that what sets all human beings apart is their equal capacity for moral agency, and that their equal status and dignity as ends in themselves is founded on that capacity.

Kant's moral philosophy brings this idea into unprecedentedly sharp focus. Equal capacity for autonomy is not a question of equal technical or computational skills, or equal experience. What Kant asserts is the equal *normative insight* of all humans, and on that basis their equal positive freedom. This proudly humanist doctrine is protected by Kant's transcendentalism about rational agency. Human beings are set apart from nature by this power. In principle, it seems, Kant's doctrine is that every rational being, through its own inherent spontaneity—through the causality of freedom—has insight into all pure reason relations. Quite clearly, however, what especially matters to Kant is an egalitarianism of pure *moral* insight. It drives Kant's insistence that morality is accessible to everyone, that arriving personally at a correct moral judgment is open to the simplest human being. Every human being merits absolute respect, because every human being has absolute moral insight.

Shorn of the transcendental backing Kant provides for it, however, egalitarianism about moral autonomy is an empirical thesis (though it also assumes a cognitivist moral epistemology). It is a much stronger thesis than the claim that moral agents must be self-determining. It is true that reason relations of any kind can only be known, in the first instance, by personal insight. However, that does not entail that everyone must know them in that way. It may be that some agents, perhaps many, can

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

know, or truly believe, or just more-or-less wholeheartedly endorse reason relations, only through faith, or trust, or on authority, or from a desire to be at one with others. While the very idea of moral agency requires self-determination, it is not obvious that it requires autonomy.

If the capacity for moral insight varies, we are left with three possibilities. The first is to say that the Insight Principle does not apply to everyone, because not everyone is fully autonomous in the way Kant envisaged; people vary in their degree of moral insight. Moral norms, however, *do* apply to everyone; but some people must come to accept them in a less than fully insightful way. Those people are self-determining, but they are not autonomous.

This is a view that has been taken for granted in many societies at many times. The notion that God or the gods tell us the moral law is a version of it. The gods usually have authoritative priestly interpreters; however, if this is to be morality rather than mere command, there must be a spontaneous disposition to accept the moral law implanted in people's hearts. But today even this degree of hierarchy sticks in many people's gullets. There are then two alternative options. Of these, the first, historically very important over the last century or so, subverts the notion of autonomy from within. From the egalitarianism of autonomy it retains only the negative claim that no one is better placed to judge of fundamental normative claims than anyone else is. This, it says, is because we cannot justify the degree of objectivity in moral judgment that talk of greater or less moral insight assumes. This view agrees that your moral obligations are those given by your own judgment of conscience; but it does not claim that conscience is a faculty of knowledge. *Authenticity* of acceptance still matters, but not because authentic as against inauthentic acceptance delivers moral truth. Instead, authenticity, understood as deciding sincerely or "authentically" for oneself as against being led by others, becomes a peculiar, legitimacy-conferring, modern virtue of the person, though bereft of any epistemological foundation. We "respect" the authentic person's moral judgment not because it is "right" but because it is *his* sincere judgment. This is still a strong ideal of self-governance, but it drops autonomy and with it moral objectivity.

The second alternative option preserves autonomy, the Insight Principle, and the unity and objectivity of morality. But it holds that those who lack insight into some moral norm do not have the moral obligations associated with that norm. For—contraposing the Insight Principle—if they cannot tell for themselves that they have a certain moral obligation, then they do not have it. In a specious sense, this point preserves a thesis of equal autonomy: everyone is equally capable of recognizing what reasons and moral obligations there are *for him.* However, not everyone is equally capable of telling what reasons and moral obligations there would be for him if only he were capable of the necessary insight: in that more obvious sense, not everyone is equally autonomous. Furthermore, on this

view, not everyone has the same moral obligations. On this view, the standards of moral obligation are higher for the more insightful (and thus more autonomous) than for the less insightful. Thus, while this view preserves autonomy and moral objectivity, it gives up egalitarianism about autonomy: it is the standpoint of moral autonomy *without* egalitarianism.

We can trace the implications of this standpoint for modern moral assumptions further, by bringing together the theses discussed in this and the previous section: the Insight Principle and the categoricity of morality.

VI. Autonomy and the Insight Principle

Morality is linked to reasons for action by the crucial implications stated in thesis (7) and thesis (9) in Section IV. Those implications can be contraposed. If the agent has no warranted reason not to α, then it is not morally wrong for him to α. Thus, a substantive theory of practical reasons can shape one's views of moral obligation. This obviously applies, for example, to egoistic and instrumentalist theories of practical reasons.[17]

A distinctive theory of practical reasons is also a main source of Bernard Williams's critique of morality—as he notes, for example, in the following passage:

> Blame rests, in part, on a fiction; the idea that ethical reasons, in particular the special kind of ethical reasons that are obligations, must, really, be available to the blamed agent. . . . *He ought to have done it,* as moral blame uses that phrase, implies *there was reason for him to have done it,* and this certainly intends more than the thought that we had a reason to want him to do it. It hopes to say, rather, that he had a reason to do it. But this may well be untrue: it was not in fact a reason for him, or at least not enough of a reason. Under this fiction, a continuous attempt is made to recruit people into a deliberative community that shares ethical reasons. . . . But the device [i.e., the "fiction"] can do this only because it is understood not as a device, but as connected with justification and with reasons that the agent might have had; and it can be understood in this way only because, much of the time, it is indeed connected with those things.[18]

In saying "*He ought to have done it,* as moral blame uses that phrase, implies *there was reason for him to have done it,*" Williams is assenting to

[17] Egoism holds that what there is reason for you to do wholly depends on what promotes your own good. Instrumentalism holds that what there is reason for you to do wholly depends on the objectives you actually have.

[18] Bernard Williams, "How Free Does the Will Need to Be?" in Bernard Williams, *Making Sense of Humanity* (Cambridge: Cambridge University Press, 1995), 16; emphasis in the original.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

something like thesis (7), or indeed thesis (9).[19] Thus, he accepts that the categoricity of the moral is part of our concept of the moral. He then adds his theory of internal reasons, which says that there is reason for *x* to α only if α-ing will serve a motive that *x* has, and he notes the contrapositive implication. If α-ing *doesn't* serve a motive that *x* has, it won't be true that *x* has a reason to α, and thus it won't be true that *x* ought to have α-ed, "as moral blame uses that phrase"; that is, it won't be true that *x* had a moral obligation to α. The fiction we are then led into, Williams says, is that of treating *x* as though he really *did* have the relevant reasons, the reasons which generate a moral obligation.

We are led into this fiction, he thinks, because we sometimes want to blame people who, by his internalist standards, simply do not have these ethical reasons. We want everyone equally to fall within the scope of blame. But suppose there are people who just don't have certain reasons which we regard as morally salient, because these reasons have no grounding in their motives. In that case, we fantasize that they do have these reasons after all; we imagine that there are reasons that apply to them irrespective of their motives.

I have argued elsewhere that Williams is mistaken in tying reasons to motives.[20] For present purposes, however, we can waive that issue; the relevant point is that an analogous internalist argument can be mounted from the Insight Principle alone. Even if what reasons there are for you does not depend on your *motives,* given the Insight Principle it still depends on something about you: on what reason relations you can grasp by personal insight. If moral agency is autonomous agency, what makes one a moral agent is the capacity to recognize morally salient reasons as such, in the first-person way. Williams presents a motivational form of internalism; the Insight Principle presents a cognitive form of internalism.[21]

Similar issues thus arise. According to the Insight Principle, if a moral agent has a moral obligation, he is capable of personal insight into that obligation. Conversely, then, if an agent cannot know by personal insight that he has an obligation, then he does not have it. Furthermore, to know in the first-person way that he is subject to a moral obligation, he must be able to recognize, in that way, not only the relevant practical reasons but also that failure to act in accord with them will give sufficient reason for guilt and others' blame. A lack of personal insight

[19] I say he is assenting to something like these, because he does not discuss whether moral obligation is connected to reasons as such or to warranted reasons.

[20] John Skorupski, "Internal Reasons and the Scope of Blame," in Alan Thomas, ed., *Bernard Williams* (Cambridge: Cambridge University Press, 2007), 73–103.

[21] Williams treats Kant's view as a "limiting case" of internalism (Williams, "Internal Reasons and the Obscurity of Blame," in Williams, *Making Sense of Humanity,* 44, note 3); see also Bernard Williams, "Replies," in J. E. J. Altham and Ross Harrison, eds., *World, Mind, and Ethics: Essays on the Ethical Philosophy of Bernard Williams* (Cambridge: Cambridge University Press, 1995), 220, note 3. I do not think it is quite right to treat Kant as a limiting case of *motivational* internalism, but he certainly is a *cognitive* internalist.

at either point will remove the agent from the scope of answerability and blame.

Thus, the Insight Principle has the implications for morality and answerability that interested Williams. If, as he suggests, our drive is to bring everyone within the scope of answerability, then we must either insist that people have reasons even if they have no insight into them, or insist that everyone has insight into reasons, at least the reasons involved in moral agency. The big difference is that the Insight Principle, unlike Williams's theory of internal reasons, is indisputably a major influence on modern morality, in the shape of characteristically modern doctrines of conscience. No modern theorist of morality as autonomy can set the Insight Principle aside.[22] So that leaves only the second option, that of insisting that everyone has equal, full insight into moral obligations and their underlying reasons.

I doubt whether commitment to equal moral autonomy is solely a result of the wish to make everyone equally answerable, that is, liable to blame. Equal moral autonomy is a modern ideal, closely associated with notions of equal worth, and thus with ethical foundations for liberal democracy that are both philosophically problematic and deeply held. That may explain the fierceness with which people cling to Kant, who surrounds these particular foundations with a formidable *a priori* defensive wall. There is the transcendental armor-plating of the causality of freedom, the idea that the Categorical Imperative can be deduced from autonomy, and the further idea that it provides a criterion of moral obligation which "the most common and unpractised understanding"[23] can apply without difficulty.

In reality, the capacity for moral insight is an empirical capacity. It calls—among other things—on emotional intelligence, the ability to judge evaluative reasons. There is no *a priori* guarantee that it is distributed equally. Nor is it, I have argued, a conceptual truth that moral agency requires a capacity for full first-person moral insight. What it requires is the capacity for self-determination.

How important is this? It is important in that it should shift our understanding of moral judgment and moral agency away from the Kantian model. Yet it need not shift it further, so to speak, than to Hegel's critique

[22] J. B. Schneewind, *The Invention of Autonomy* (Cambridge: Cambridge University Press, 1998), recounts how a command model of morality increasingly gave way to a model of morality as self-governance, with Kant's notion of autonomy as a culmination of the process. The striking subsequent history of autonomy remains to be told. (Schneewind takes it that autonomy is implicit in the notion of 'self-governance.' My use of 'self-governance' is somewhat different: self-determination is the basic kind of self-governance required for any moral or rational agency; autonomy is stronger. A *pure* command model would not be morality: it must combine with some notion of inner assent.)

[23] Immanuel Kant, *Critique of Practical Reason,* in Kant, *Practical Philosophy,* ed. and trans. Mary J. Gregor (Cambridge: Cambridge University Press, 1996), 169 (Akademie edition, 5:36).

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press

of *Moralität* (by which he means the standpoint of abstract and individualist conscience, of which Hegel takes Kant's moral theory to be an outstanding example) and Hegel's account of modern *Sittlichkeit* (the broadly accepted moral convictions and practices of a society).[24]

Self-determination, as Hegel profoundly sees, requires the context of moral community—*Sittlichkeit*. It remains true, however, that between self-determination and autonomy lies not a chasm but a territory with no clear-cut frontier. In particular, moral principles are not like positive law. To determine what to do in accordance with positive law, or external command, I need to know what that law is. Furthermore, it is an open question whether I have reason to do what that law requires me to do. Moral self-determination is not like that. I can know my moral obligations by reflection on reason relations alone.

Thus, moral self-determination requires that I freely internalize reasons *as* reasons—that I accept or assent to them *as* reasons and be capable of acting on them. On Hegel's view, they should "not merely lay their claim on" me "as external laws and precepts of authority to be obeyed, but have their assent, recognition, or even justification in [my] heart, sentiment, conscience, intelligence, etc."[25]

Moral agency requires at least this much "making them one's own." Questions remain about the moral psychology of normative assent, about what elements of thought and feeling ("heart, sentiment, conscience, intelligence") can be involved. We shall still need something like the Insight Principle, but our way of answering these questions will determine how we broaden it.

This, in turn, raises issues about the epistemology of moral judgment. Its epistemic bases are, on the one hand, one's own spontaneous dispositions of feeling and will, and, on the other, recognition of and critical response to the spontaneous dispositions of others. This involves the influence of others, which works through much more than explicit discussion, let alone argument. Still, explicit discussion, including argument, is essential for solid normative knowledge. If there is to be knowledge of reason relations, it must come from pure insight into reason relations on the part of at least some people. However, that does not imply a one-way transmission from these knowers to others. Normative knowledge is inherently social. It rests on discussion with others as well as personal spontaneity—discussion from which no voice is excluded. The point applies to all normative knowledge, but especially to moral knowledge, because of its sentimental and practical bases. The resulting picture of *Sittlichkeit,* moral life, actually provides (to my mind) a better foundation for democratic ideals than a picture of ourselves as

[24] Hegel's most extended discussion of this contrast is to be found in his *Elements of the Philosophy of Right* (1821). For commentary, see (for example) Allen W. Wood, *Hegel's Ethical Thought* (Cambridge: Cambridge University Press, 1990).

[25] G. W. F. Hegel, *Encyclopedia of the Philosophical Sciences* (1817), sec. 503.

perfectly autonomous knowers, each equipped with access to a fail-safe procedure for deducing moral conclusions, which we operate in the solitude of conscience.[26]

VII. Conclusion

As I said at the beginning, two deep features of moral judgment should guide our discussion of morality: the first is that moral judgment is rooted in the rationality of the feelings; the second is that moral agency is self-governing agency.

What I hope to have shown is that moral judgment is judgment about evaluative reasons, specifically about reasons to blame. The sentiment of blame can be characterized in a noncircular way by its objects and practical dispositions. Furthermore, it is by virtue of its object (failure to respond to certain reasons highlighted by the sentiment itself) that blame gives rise to the categoricity of moral obligation.

Because moral agency is self-determining agency, I must be able to tell, by reflection on my actual epistemic state, what my moral obligations are, just as I must be able to tell what my warrants are. So moral obligation, like warrant, is relative to epistemic states. As to self-governance, it can be less than autonomy but must be more than authenticity. Yet these latter conceptions of moral agency have a strong influence on our ethical culture. A more realistic conception of self-determining moral agency is something that we greatly need.

Finally, though I have not developed this argument, I believe that the bases of normative knowledge involve an interplay between individual spontaneity and communal discussion: in the case of morality, between individual conscience and the moral sentiments underlying common moral sense. Moral reflection is the "common pursuit of true judgment," and self-governance, however individually insightful, can only count as knowledge, and work properly, within the context of that common pursuit.

Moral Philosophy, University of St. Andrews

[26] It would be a one-sided caricature to say that Kant himself propounded this picture, since there is much in his ethics taken as a whole that goes against it. Yet it is also true that Kantian ethics has historically both played to and entrenched the picture.

https://doi.org/10.1017/S0265052509990197 Published online by Cambridge University Press