

## Internalism about Responsibility

By R. Jay Wallace  
University of California, Berkeley

**Abstract:** Internalism in ethical theory is usually understood as the view that there is a non-contingent connection between moral rightness and motivation. In this paper, I develop the analogous suggestion that there is a non-contingent connection between moral rightness and responsibility.

Externalists about motivation hold that it might be an open question whether people have reason to comply with moral principles. Externalists about responsibility similarly hold that it might be an open question whether there is reason to blame people for failing to comply with moral principles. From the fact that an action would be wrong, for instance, it doesn't follow that our social relations should be structured in a way that blames people when they perform actions of that type. That an action is wrong is not itself a reason for holding people accountable on the externalist view.

I argue against this approach, exploiting the analogy with internalism about motivation in support of an internalist account of responsibility. In particular, I suggest that moral principles have it as part of their function to ground responsibility relations between people, and that an interpretation of those principles that was ill-suited to this function would be deficient on that account. Morality, on the modern conception of it that I shall be sketching, is an essentially relational or social phenomenon, and internalism about responsibility is one aspect of its relational character.

There is a traditional debate in ethical theory about the relation between moral rightness and motivation. Internalists, as they are sometimes called, hold that there is a non-accidental connection between these things. According to this position, sincere judgments about what is morally required are necessarily motivating (or necessarily motivating insofar as the person who makes them is rational, or reasoning correctly).

Externalists deny this claim, maintaining that one can fail to be motivated in accordance with moral thought, and that such failures do not necessarily entail irrationality or other departures from correctness in reasoning. One thing that is at stake in this debate is the question of the constraints to which philosophical accounts of morality are answerable.

If internalism is correct, then a theory of morality should aim to make sense of the non-contingent connection that is postulated between moral thought and motivation. A theory that fails to meet this constraint will be *prima facie* inadequate, and its acceptance

would involve a revisionist understanding of the phenomenon that it attempts to account for.

In this paper I want to explore an analogous question about the relation between morality and responsibility: is moral rightness necessarily connected to responsibility relations between people, or is the connection between these things merely contingent? Internalism about responsibility, as I shall understand it, is the view that rightness and responsibility are non-contingently connected. The principles that determine what it is morally permissible to do must also be suited to structure and to ground relations of responsibility among the members of a moral community. My aim in the present paper is to offer a sympathetic statement of this position, exploiting the analogy with internalism about motivation to mount a modest defense of it.

The discussion divides into four parts. In the first, I offer a brief account of internalism about motivation, explaining its rationale on the interpretation that I favor. In part two, I draw on my discussion of internalism about motivation to develop an analogous version of internalism about responsibility. In part three I refine my statement of internalism about responsibility, focusing on the issue of the significance of agents' attitudes for our responsibility relations. In the brief concluding section I turn to the implications of internalism for our understanding of the content of moral principles, showing how the internalist approach to morality leads to a relational interpretation of its requirements.

### **1. Rightness and Motivation**

Traditional internalists postulate a non-contingent connection of some kind between morality and motivation. But how exactly should the connection they envisage be understood? The most common version of the position, which has been called "moral judgment internalism"<sup>1</sup>, is cashed out in psychological terms. The idea here is that sincere endorsement of a moral judgment carries with it some tendency to be motivated

to comply with the judgment. According to this view, there is a distinctive psychological condition that must be satisfied if an agent is to count as accepting the claim expressed by a moral judgment.<sup>2</sup>

Moral judgment internalism has figured prominently in metaethical discussions during the past century. It has been appealed to by expressivists, for instance, who claim that moral judgments are not the kind of cognitive attitudes that aim to represent the way the world is, but involve instead desires or pro-attitudes of some kind, which moral assertions express or give voice to.<sup>3</sup> They argue that moral judgments are necessarily motivating, and that we can make sense of this dimension of moral thought only if we understand those judgments to express the kind of non-cognitive attitudes that are essentially involved in motivations to act. Indeed, it seems fair to say that moral judgment internalism in some form is the primary consideration alleged to support expressivist approaches—it is the expressivist's Ur-argument, if you will. But the thesis is notoriously problematic. The basic difficulty with it is that it is very hard to see how one could defend moral judgment internalism against possible counterexamples to it in a way that does not simply beg the question.

Traditionally, opponents of moral judgment internalism invoke the figure of the skeptic about morality.<sup>4</sup> This is someone who understands moral language well enough, but who doubts or rejects the significance of moral properties and distinctions in reflecting about how to act. The stance of the moral skeptic seems a coherent one, insofar as we can imagine a person who is competent at moral discourse, but who questions its significance for their own practical reflection. Indeed, we can imagine this happening in our own case, envisaging a trajectory that takes as its starting point our present commitment to moral ends, and arrives at the position of the skeptic. In this scenario, we would retain our actual competence with moral predicates and distinctions, but lose any concern to comply with moral requirements, having come to doubt the significance of those requirements for our deliberation about what to do. The prospect

of this imaginary trajectory in our thinking about morality will probably seem horrifying from our present point of view, and the skepticism in which it terminates may in fact be substantively mistaken. But it appears to be a perfectly intelligible scenario, and yet moral judgment internalism appears to rule it out, on grounds that seem questionably *a priori*.

We should therefore reject moral judgment internalism. But there is a deeper insight in this position that we should also try to hang onto. What is attractive about internalism, I believe, is the thought that moral considerations at least purport to have normative significance. They present themselves to us as reasons for action, in the basic normative sense of being considerations that count for or against courses of action that are open to us.<sup>5</sup> Thus it is not merely a brute fact about us that we tend to find ourselves drawn to actions that we judge to be morally right or valuable. From the first-person point of view, these moral characteristics strike us as considerations that recommend or speak in favor of the actions to which they apply. Furthermore, the fact that they strike us as normative in this way is connected to our tendency to be motivated in accordance with the moral judgments that we endorse. Here is one way we might develop this insight into an account of the relation between moral properties, moral judgments, and motivation.

Assume, first, that it is a condition of rationality that agents are motivated in accordance with the normative judgments that they sincerely endorse. It is presumably not impossible to fail to be motivated in accordance with such judgments—something like this happens, for instance, in cases of weakness of will and in some forms of self-deception. But when such cases arise, it is natural to say that the agents involved in them are irrational, insofar they fail to be motivated as they themselves judge that they ought to be. It is thus part of being a rational agent to have dispositions to action (and thought) that are in accordance with the normative reasons one acknowledges to obtain.

Assume, second, that moral considerations do in fact represent genuine reasons for action, having the status of considerations that, for any agent, count for or against that agent's acting in certain specified ways. If this is the case, then we can say that agents are necessarily motivated to act in accordance with moral requirements, to the extent they are deliberating correctly and are otherwise practically rational. The condition of correct deliberation rules out cases in which an agent does not acknowledge the truth of moral judgments, or does not acknowledge that such judgments have normative significance for practical reflection. And the rationality condition rules out agents whose motivations fail to align with their own verdicts about what there is reason to do, in the style of weakness of will.

The combination of these two conditions—which I shall henceforth refer to as internalism about motivation—provides a plausible (if rough) characterization of the action-guiding dimension of morality. The requirement combines a claim about the effects of a certain class of judgments, namely normative judgments about what there is reason for one to do, with a substantive thesis about morality, to the effect that it is itself a source of reasons in this normative sense.

The position that results from this combination of claims leaves room for the kind of moral skepticism that is excluded *a priori* by moral judgment internalism. Indeed, there are two points at which space might open up between an agent's sincere moral judgments and their motivations to action. First, a person might accept both the truth of some moral judgment, and the normative significance of the judgment thus arrived at, without intending to act accordingly. We might believe, for instance, that we are morally obligated to help victims of political persecution in our community, and that our being so obligated speaks strongly in favor of acting accordingly, without really caring about whether we ourselves succeed in providing such assistance when we are in a position to do so. In this scenario, we fail to be motivated in accordance with normative claims that we ourselves accept. Insofar as we accept the normative authority

of moral principles for practical deliberation, however, it would not be very plausible to describe us as skeptics about morality. We will rather be acting in ways that are questionable or misguided by our own lights, a condition that involves irrational weakness of will (something along the lines of fecklessness or depression) rather than moral skepticism strictly speaking.

In a different scenario, we might acknowledge the truth of moral judgments without yet accepting that conclusions about what is morally right and wrong have any normative significance at all for us (or perhaps for any agent). Thus we might accept that we are morally obligated to help the locally oppressed, without granting that this fact by itself counts in favor of our doing anything to provide such assistance when we can. This in fact seems to describe much better the outlook of a moral skeptic. What skeptics typically doubt or challenge is the normative significance of morality—the idea (for instance) that one ought to help the politically persecuted in one’s community, just because and insofar as the failure to do so would be wrong.

While allowing for these possibilities, however, the motivation condition, as I have formulated it, still gives expression to the idea that there is a non-contingent connection between morality and motivation. At its heart is the thesis that rightness and other moral considerations represent reasons for people to act in accordance with them. This thesis is a version of the position that is sometimes called “rationalism” in ethics; if it is true, then we can say that those agents who fail to be motivated to act in accordance with their moral judgments are making a substantial mistake of some kind. Either they are failing to acknowledge the normative significance of facts about moral rightness and permissibility, insofar as they deny that people have reason to act only in ways that are morally permissible. Or they grant this normative dimension of morality, but irrationally fail to do what they themselves acknowledge that they have reason to do. Conversely, when we form intentions that are in compliance with moral principles, our being motivated in this way is not a mere optional extra, something that just happens to

be true of us as a contingent matter of psychological fact. Rather, it is a response that is rendered appropriate by the nature of the moral considerations that we judge to obtain, insofar as those considerations do in fact constitute reasons for acting in accordance with them. They are considerations that merit our compliance with them, and it is in that sense that moral motivation is non-contingently connected to the moral principles it is responsive to.

But what is the connection between rationalism, in the sense at issue here, and moral concepts? Does it follow *a priori* that if X is the right thing for me to do, then X is something that I have a very strong reason to do? Or is the normative standing of moral considerations a merely contingent fact about them, one that isn't guaranteed by correct application of the concepts involved in moral judgment? The issues here are delicate ones.

On the one hand, it is certainly part of our conventional understanding of morality that it constitutes a domain of reasons for action. Thus we typically cite moral considerations in discussion with other people, as factors that are of direct normative significance, counting for or against options that are under active consideration. We raise our children to treat moral considerations in this way, for example. Furthermore, many of us structure our (adult) deliberations on the supposition that moral considerations have normative standing, taking facts about rightness and moral value to have direct significance for our decisions about what to do. These considerations suggest to me that it is one of the familiar platitudes about morality that its central concepts (such as rightness and permissibility) are imbued with normative significance. On the other hand, there has to be room for the skeptical position that coherently questions whether people really do have reason to comply with the standards that define what is morally right and wrong. The skeptical position might be mistaken as a matter of fact, but it isn't merely confused; as I suggested above, one can grant that it

would be wrong to do X, and yet without contradiction deny that this is a reason against acting in that way.

The best way to do justice to these twin pressures, it seems to me, is to take an element of revisionism to be endemic to the skeptical position. Skeptics, insofar as they deny that moral rightness is reason-giving, are denying one of the platitudes that help to fix the meaning of the concept of the morally right. They are thus denying that there is anything in the world that completely answers to this moral concept. It doesn't follow, however, that the position they are adopting is merely confused or incoherent. In saying that people don't have reason to comply with the standards of moral rightness, they can be interpreted as suggesting that the properties in the world that most closely approximate to our concept of the morally right are not properties that have normative significance. Morality cannot, as a result, be everything that it represents itself as being, insofar as one of the platitudes that help to fix the concept turns out to be false as a matter of fact. But this strikes me as a plausible thing to say about the kind of skepticism I have been considering. It is a modestly revisionist position, denying something that strikes us as partly constitutive of the moral basic concepts in the first place, namely the direct significance of the properties they describe for deliberative reflection about what we are to do.

This brings out a second respect in which the connection between morality and motivation is non-contingent, on the internalist position I have been sketching. Given the rationalist thesis that moral considerations are reasons for action, the motivation to comply with them is a response that is appropriate to its proper object. But this thesis itself is non-contingently connected to the central moral concepts, in such a way that an account of morality that denies the thesis will therefore be at least modestly revisionary.

## 2. Rightness and Responsibility.

Let's now turn to responsibility, considering how an internalist position might look that is modeled on the view sketched in the preceding section.

The first thing to note is that it is not plausible to suppose that attributions of responsibility are built into the act of moral judgment. The judgments that one might expect most closely to involve such attributions are judgments that acknowledge moral shortcomings, including above all judgments to the effect that an agent has acted wrongly or impermissibly. These are the kinds of things that people are typically blamed for, and internalism about responsibility might accordingly be understood to hold that judgments of wrongdoing amount to acts of blame. But this form of "moral judgment internalism" seems implausible, and for reasons that are similar to the reasons that speak against moral judgment internalism about motivation. One can sincerely believe that an agent has acted wrongly without blaming the agent on that account, and this defeats the suggestion that there is an *a priori* necessary connection between judgments of moral wrongdoing and blame.

Appreciation of this point will be enhanced by brief reflection on the nature of blame. Philosophers have offered a variety of conflicting accounts of moral blame, and there isn't space here to discuss their merits in any detail. So I will cut to the chase and simply assert that on the view I find most attractive, blame should be understood in terms of the reactive emotions of resentment, indignation, and guilt.<sup>6</sup> To blame a person for something, on this view, is to think or judge that the person has done something morally impermissible, and to be subject on that account to an appropriate emotion from this class that is directed toward the wrongdoer. If the agent has done something to wrong me in particular, for instance, then I will react with blame when I resent the agent for having treated me in this way. Understood in these affective terms, blame is a matter of one's emotional responses to lapses from the standards defined by moral requirements. In particular, it is a way of being exercised by such lapses that shows that

one has internalized a concern for moral values. One cares about those values, where this in turn involves a characteristic form of emotional vulnerability to offenses against them. Blame can be understood as a manifestation of one's attachment to morality; it is a reaction that reveals that it matters to one whether people succeed in complying with moral requirements in their interactions with each other.<sup>7</sup>

If these brief remarks are on roughly the right lines, however, then it should hardly be surprising that judgments of moral wrongdoing can come apart from either blame or its expression. To judge that a person has acted wrongly is to judge that the person has fallen short by reference to the standards that determine what it is morally permissible to do. But it is an elementary feature of human psychology that we are able to make judgments of this kind even if we do not particularly care about the standards that are at issue in the judgments. We can acknowledge that moral ends have been flouted, for instance, without particularly valuing the forms of relationship that are promoted and made possible by pursuit of those ends. Under these conditions, we will not be prone to blame in the cases in which we acknowledge wrongdoing to have occurred, insofar as we will not have internalized the concern for moral values that is a precondition for emotional reactions of this distinctive kind.

As we saw in the preceding section, however, internalism about motivation is not in any case plausibly understood to involve a necessary psychological connection between moral judgment and the reactions with which it is taken to be non-contingently connected. We do better, I suggested, to think of this connection in normative terms, taking it to rest on the standing of moral considerations as reasons for the motivational responses with which it is non-contingently linked. Let us now turn to the question of how this model might be applied to the case of internalism about responsibility.

There are, I believe, two respects in which we might plausibly understand moral wrongness to be connected normatively to our responsibility reactions. First, the values around which morality is organized might be understood to be values that people in

general have good reason to internalize and to care about. These stances involve, as we have seen, forms of characteristic emotional vulnerability, including in particular a tendency to experience the reactive emotions in cases in which people act in ways that flout moral requirements. Even if it is true that we can acknowledge cases of moral wrongdoing while failing to internalize in this way a concern for moral ends, it might still be the case that we in general have reason to adopt this distinctive affective stance, developing the kind of concern for moral ends that would render us susceptible to the reactive sentiments when those ends are offended against. And indeed this seems to be the view that many of us implicitly adopt in our practices of moral education and habituation. We don't think that it is optional for our children to be brought up to care about moral values, but seem to view those values as ones that everyone has good reason to internalize. Our interactions with children are accordingly designed in part to inculcate in them an emotional commitment to morality, of the kind that will leave them systematically vulnerable to negative reactions of blame and opprobrium when people wantonly disregard moral requirements in their interactions with others.

Morality in this way seems to differ from many other domains of value. There are plenty of things that we acknowledge to be genuine goods, without taking it to be important that all members of the younger generation should be brought up to internalize a specific concern for them. Philosophy, for instance, is an exceptionally worthwhile activity (in my humble opinion). And yet it would hardly be a failing on our part if our children should fail to develop an emotional commitment to this particular activity, coming instead to care about (say) physics or cabinetmaking. Morality seems different in this respect, constituting a domain of value that we all have reason to become emotionally invested in. We might be wrong to think this, of course, but if so it would come as a shock and a surprise, overturning an assumption that is central to our understanding of morality.<sup>8</sup>

So this is one way in which morality might plausibly be understood to stand in normative relation to responsibility. A second normative connection that might plausibly be taken to obtain links moral facts about human actions to specific attributions of responsibility for those actions. Thus, the fact that A has deliberately done something wrong is at least a defeasible reason for those who have been wronged to resent A for having treated them in this way. It makes sense to react in this way to acts of this kind, and such reactions are therefore rendered appropriate or even called for by the fact that the actions to which they are directed were morally impermissible.

In saying this, I mean in part to be alluding to the fact that blame reactions would not be fair or fitting in the complete absence of wrongdoing on the part of the agent at whom they are directed. Deliberate wrongdoing on the agent's part renders blame appropriate, insofar as this reaction is not objectionable in the way it would be if the agent's behavior had been beyond moral reproach.<sup>9</sup> But blame seems called for by wrongdoing in a somewhat stronger sense as well: given the general reasons we have to internalize the kind of concern for moral ends that renders us vulnerable to reactive sentiments in the first place, those emotions are positively appropriate responses to actions that are wantonly impermissible. A failure to experience them when e.g. someone has wronged us would tend to indicate the absence of the sort of emotional investment in moral values that we generally take to be desirable. By the same token, the fact that A has complied with moral requirements in A's interactions with other people provides a certain level of normative protection from the emotional reactions that are characteristic of blame. If A's treatment of other people does not involve any moral wrongdoing, then it would ordinarily not be appropriate for them to react to A's actions with the kind of opprobrium and focused hostility that are involved in blame.

Of course, these specific connections between the moral features of an agent's actions and the reactions of other people may often fail to obtain as a matter of fact. We might for instance be strangely indifferent to the fact that A has wronged us, or

alternatively feel resentful of A despite the fact that A has not done anything that is genuinely impermissible. But these reactions would often be subject to normative objections of various kinds: e.g. as too mild, or as unduly harsh and, well, blaming, objections that imply that there are good reasons for modulating one's reactions to what A has done. This is the sense in which we might take there to be a normative connection between the moral qualities of actions and the specific reactions that those actions might evoke on the part of people variously affected by them.

Any plausible development of this position will require significant qualification. There may be normative connections between the moral qualities of actions and our responsibility-involving reactions to them, but the connections are capable of being defeated or overridden in particular cases. Thus, many cases of wrongdoing are too remote from my own life and experience for it to be plausible to think that I have good reason on balance to become emotionally exercised about them. Particularly if the wrongdoing that is at issue is not egregious, and the individuals involved are unknown to me, it might be meddlesome or sanctimonious of me to react to the wrongdoing with indignation. The fact that the victim was wronged would perhaps give us some reason to blame the agent, but the reasons are outweighed on the other side by a variety of considerations that count against reacting in this way.

Another set of cases of this general kind might be those that involve acts of wrongdoing by agents whose powers of moral competence and control are seriously impaired. Even when such agents have wronged me in particular, the fact that they lack the general capacities for moral understanding and control would ground a strong moral objection to my reacting with the standard feelings of blame, so that blame is not a reaction that is really called for or appropriate under the circumstances.<sup>10</sup> Here again, we might think of facts about wrongdoing as providing reasons for opprobrium that are defeated by the reasons against reactions of this kind, given other relevant facts about the agent of the wrongdoing in the case at hand.

I have so far described two respects in which we might plausibly understand morality and responsibility to be connected to each other normatively. These connections, assuming them to obtain, give a sense to the suggestion that morality and responsibility are related in a way that is not merely accidental or contingent. Holding people responsible isn't a purely optional stance that we might or might not adopt toward them, as we happen to see fit. Rather, it is a response that is inherently answerable to facts about the moral qualities of people's actions, and this in two distinct respects. First, the values around which morality is organized are ones that we all have reason to internalize and to care about, so that we become susceptible to the emotional reactions constitutive of blame on occasions when those values are thwarted or defeated. Second, the moral qualities of individual actions give us defeasible reasons both for blaming their agents and for refraining from blaming them, depending on the qualities that are instantiated in the actions they perform.

There is a third and final point to make as well. In discussing internalism about motivation, I suggested that it might in part be taken to involve a thesis about the concept of morality. The thesis is to the effect that it is part of our understanding of the concept of morality that moral considerations, such as rightness or permissibility, are reasons for action. This thesis might turn out to be false, but if so that result would involve a modest revisionism about morality, the concession that nothing in the world fully matches the contours of our concepts of moral rightness and permissibility. I now want to formulate a similar thesis about the relation between moral concepts and responsibility.

The basic idea is that it might plausibly be taken to be one of our platitudes about morality that moral standards regulate responsibility relations between people. Thus, the standards that determine what it is right or permissible to do have it as part of their function to be internalized emotionally, in the way that renders us susceptible to blame reactions when those standards are violated. On this view, it is part of our concept of

morality that e.g. moral wrongness is not only a reason for action (a reason, specifically, to avoid actions that have the property), but also a reason for responsibility reactions (a reason, specifically, for blaming the agent who performs an action with this property). An important consideration in support of this interpretation of the concept of morality is the fact that we tend to identify a society's moral standards by looking to the requirements that are implicit in the reactive sentiments of its members. Our collective moral standards, on this plausible way of seeing things, just are (in part) the standards whose deliberate violation attracts the emotional opprobrium of blame on the part of the members of our community.

In saying this, of course, I don't mean to be saying that we are infallible about what morality requires of us. We might attach opprobrium to the wrong things, blaming people when they engage in eccentric sexual practices with other consenting adults, for instance, and failing to blame them when they turn their backs on the basic human needs of vulnerable members our community. We could express this possibility by saying that our moral standards are mistaken or misguided in cases of this kind, and this tells us something about the concept of morality: namely, that we think of morality as a set of standards that function to regulate our responsibility relations with each other, giving people reasons for reactive sentiments when they are violated.

Note, too, that this conceptual point about morality should not be taken to entail that moral standards are in fact reasons for such responsibility reactions. There is room for a skeptical position about responsibility that is analogous to the kind of moral skepticism discussed in the preceding section. This view would hold that the violation of moral standards is not after all something that we have good reason to respond to with the negative reactive emotions. If I am right, then this skeptical position, just like the corresponding skepticism about moral reasons discussed above, would involve some element of revisionism about morality. It would maintain, for instance, that our concepts of moral rightness or permissibility are not fully realized in the world as we

find it, precisely insofar as the standards that determine rightness and permissibility do not appropriately regulate our responsibility relations with each other. Our feeling that this outcome would involve a degree of revisionism is a reflection of the basic idea that a normative connection to responsibility relations is built into our concepts of moral rightness and permissibility.

### **3. Refining the Position.**

Internalism about responsibility has considerable appeal, but it cannot be accepted in precisely the form in which it has so far been stated. I shall approach the need for refinement by considering a more general issue, having to do with the relation between permissibility, responsibility, and the agent's intent.

At first glance, it seems extremely natural to suppose that the moral permissibility of our actions might at least sometimes depend on the intentions with which they are performed, including in particular the reasons for which we decide to carry them out. Thus, it appears to be wrong to fire an employee because the employee has declined your sexual advances, even if it would be permissible to fire the same employee for any of a range of other reasons that might be available in the case. Similarly, refusing to rent an apartment to someone on account of their race seems to be wrong, even when one has wide discretion to decide for oneself whom one will enter into a contractual agreement of this kind with. Debates about just war and the ethics of abortion have also frequently invoked principles (such as the doctrine of double effect) that make the permissibility of actions that bring about harm depend in part on the agent's reasons for choosing to perform them.

But these appearances might be questioned. T. M. Scanlon, for instance, has recently mounted an interesting argument for the conclusion that the permissibility of actions almost never depends on the intentions with which agents carry them out.<sup>11</sup> His case for this conclusion rests in part on ingenious interpretations of the moral principles

that determine permissibility for the problem situations, interpretations that attribute at most derivative importance to the agent's reasons for action. But intent clearly has great moral significance of some kind, and another part of Scanlon's argument is designed to locate its importance in features of moral thought that are distinct from permissibility. Specifically, Scanlon distinguishes between two dimensions of moral assessment, permissibility and meaning, contending that intent matters greatly to the latter even though it is virtually irrelevant to the former. The meaning of our actions is largely a function of the attitudes that are expressed in them, where this in turn depends on our reasons for doing what we do. Scanlon argues that responsibility relations are organized primarily around questions of meaning in this sense, rather than questions of permissibility. By acting with contempt or indifference toward someone, we impair our relationship with them, in ways that give them reasons to make the adjustments in their attitudes and behavior that are constitutive of blame, on Scanlon's account of it.<sup>12</sup> So the intentions with which actions are performed, as part of the meaning of those actions, matter greatly for questions of blame, even if they have virtually no significance for questions about the permissibility of what we have done.

This position looks on the surface to conflict with the internalist view of moral rightness and permissibility that was sketched in the preceding section. Internalism holds that moral properties of this kind have normative significance for our responsibility relations, insofar (e.g.) as we have reason to blame people when their actions are impermissible. It also holds that it is in the nature of a conceptual truth about morality that the properties it characterizes have this kind of significance, being suited to constitute responsibility relations between people. But if blame is responsive to intent, and intent is in turn largely irrelevant to questions of permissibility, it appears that the internalist position must be mistaken. Moral properties do not after all have the direct normative significance for our responsibility relations that it attributes to them.

One might respond to this challenge by questioning Scanlon's account of permissibility. But I do not wish to take that path. I believe Scanlon goes too far in denying the relevance of intent to permissibility, but he is surely right that there are many central cases in which it is not plausible to suppose that moral principles are sensitive to the meaning of the actions they regulate.<sup>13</sup> Nor do I wish to dispute Scanlon's assumption that questions of meaning are central to blame. Instead, I shall argue that the internalist position can be developed in a way that does justice to these two desiderata; indeed, once it is properly understood, internalism about responsibility can help to resolve a puzzle that Scanlon's remarks about permissibility and meaning raise for his larger moral theory.

The puzzle comes into focus when we ask the question why people should comply with the principles that determine what it is morally right to do. This is the question as to the reason-giving force of moral rightness or permissibility, which is central to the version of internalism about motivation that was presented in section 1. Scanlon himself subscribes to an internalist position of that kind, insofar as he accepts that the impermissibility of an action is a strong reason for just about anyone not to perform it. His own interpretation of this normative aspect of moral properties connects it to a valuable form of relationship between people, which he calls mutual recognition.<sup>14</sup> We relate to people in this way when we are able to justify our behavior to those who might be affected by it, by appeal to principles that it would be unreasonable for anyone to reject as a basis of general agreement. The reason-giving force of moral permissibility thus gets traced to the fact that the principles that define it are conditions for the possibility of valuable relationships of this kind.

Here's the thing about mutual recognition, however. Understood intuitively, this is a way of relating to people that is largely constituted by the attitudes we adopt toward them. To stand in this relationship with others is to acknowledge them as independent sources of claims and to regulate one's behavior accordingly, striving to act

in ways that will be justifiable specifically to them. It is thus a matter of the quality of will with which one acts. But if permissibility is independent of matters of intent of this kind, then Scanlon's account of the normative significance of this consideration for action seems to be called into question. Acting permissibly would not appear to be sufficient for mutual recognition, insofar as we can comply with principles of the moral right without having any particular concern for the standing of others as sources of claims against us. Indeed, permissibility does not even appear to be necessary for mutual recognition, insofar as we might fail to do what is objectively right even while sincerely and wholeheartedly endeavoring to act in ways that are justifiable to those affected by what we do. But if permissibility is neither necessary nor sufficient for mutual recognition, and if the value of mutual recognition is in turn the basis of our moral reasons for action, then why should we care about permissibility per se? This is the puzzle to which I alluded above.

A first step toward resolving the puzzle is to acknowledge that reasons for action are themselves reasons for intention. Thus, if the impermissibility of X-ing is a reason not to do X, then it is itself a reason to intend not to do X, because and insofar as X-ing would be impermissible. Scanlon himself holds something like this view, defending it with the observation that actions involve "judgment-sensitive attitudes", precisely on account of the states of mind of the agent that render their doings intentional performances in the first place (as opposed e.g. to mere spasms or twitches).<sup>15</sup> We are rational to the extent that we succeed in adjusting our judgment-sensitive attitudes in response to our beliefs about our reasons, and in the case of reasons for action this is a matter of forming intentions to do what we believe there is reason for us to do.

It follows from this that those agents who are responding correctly to the reason-giving force of moral properties such as permissibility and rightness will necessarily act with a certain distinctive quality of will. They will acknowledge that the impermissibility of X-ing (say) is a reason not to X, and they will form the intention not

to X for this reason. But this is the very quality of will that plausibly constitutes the necessary and sufficient subjective condition for mutual recognition. We take people to be independent sources of claims, subjects to whom justification is owed, just in case we grant the reason-giving force of permissibility and regulate our intentions in accordance with this consideration. The upshot is this. Permissibility might not itself be a condition of mutual recognition, strictly speaking. But if we postulate that it is a consideration with normative significance for human action, then agents will achieve the attitudes constitutive of mutual recognition when (and only when) they respond correctly to this consideration, acknowledging its reason-giving force and adjusting their intentions accordingly. The attitudes important to mutual recognition are thus not attitudes that make actions permissible in the first place (since by hypothesis permissibility often doesn't depend on the agent's intent at all). They are rather the attitudes one forms when one responds appropriately to permissibility as a reason for action.<sup>16</sup>

With these remarks in place, let us now return to the question of the normative significance of permissibility and rightness for responsibility. The worry about this, to put it crudely, was that blame is a response to the meaning of an agent's actions (in Scanlon's sense), and that questions of permissibility are distinct from questions of meaning. This seemed to call into question the normative connection between permissibility and blame that internalism about responsibility postulates. But the discussion of mutual recognition points the way to a resolution of this concern. We need to refine our understanding of the normative connection that is at issue, taking it to be mediated via the attitudes of the agents whose actions are up for moral assessment. Thus, to say that permissibility and rightness are normative for our responsibility reactions is to say that those reactions are properly responsive to agents' attitudes toward the moral properties in question. Those who sincerely strive to comply with the principles that determine permissibility thereby acquire normative protection from blame and opprobrium for what they do. Conversely, attitudes of indifference to these

moral considerations, or of blatant contempt for them, provide others with defeasible reasons to react to the agent with blame and opprobrium.

This is, on reflection, nothing less than we should expect, given the idea discussed in section 1 that moral considerations are in the first instance reasons for action (an idea that Scanlon himself accepts). Their having this status means that we ourselves respond correctly to such considerations when we regulate our intentions in accordance with them, striving to do what is right and to act only in ways that are morally permissible. When we internalize a concern for these reasons, one consequence will be that we as agents aim to comply directly with them, in ways that acknowledge their normative standing for our actions. But the responsibility reactions of blame and opprobrium are backward-looking responses, directed toward *agents* on account of the things they have done. In cases in which these emotional responses are at issue, our internalized concern for moral values will lead us to focus on the attitudes of the agents we are responding to, and to consider whether those attitudes reflect a due appreciation for the reason-giving force of rightness and permissibility. What we care primarily about, in this distinctive context, is not whether the agents up for assessment really acted in ways that were permissible, but whether they had the qualities of will that are constitutive of mutual recognition, responding to moral considerations in just the way we take ourselves to have reason to respond to them in the first-person perspective of deliberation. This is just what I meant in suggesting that the normativity of permissibility for responsibility is mediated via the attitudes of the agents who are up for assessment.

Consideration of the issue of intent and permissibility has thus led us to an improved formulation of internalism about responsibility. Strictly speaking, what is normative for our responsibility reactions is not the permissibility or impermissibility of what people do, but the attitudes latent in their actions toward moral considerations of this kind. Morality is suited by its nature to constitute a framework for responsibility

relations, insofar as peoples' attitudes toward it provide the normative basis for such relations.

#### **4. Conclusion: Individualistic and Relational Conceptions of Morality.**

In this brief concluding section, I want to look at the implications of internalism about responsibility for our understanding of the content of moral requirements.

Some moral theories are individualistic, focusing on the obligations that people stand under as agents considered in isolation from each other. Perfectionism (in both its Aristotelian and modern variants) is individualistic in this sense, conceiving moral (or better, ethical) standards as means to the realization of an ideal for human agents, and as deriving their significance primarily from this function. But many modern theories seem to be individualistic in the same sense. Utilitarianism, for instance, conceives of moral standards in maximizing terms, telling individuals that it is wrong to act in ways that are suboptimal in their effects on the interests of sentient beings. Whether or not a particular agent is responsive to this standard of wrongness would appear to be completely independent from the question of whether other people have reason to respond to the action with reactions characteristic of blame.

For example, utilitarians often hold that lifestyles of bourgeois consumption are morally objectionable in a world in which there are vast numbers of people living under conditions of extreme need. So long as these conditions persist, individual expenditures on consumer goods will be hard to justify in utilitarian terms, insofar as greater utility would be achieved by donating the funds at issue to an organization such as Oxfam or Doctors without Borders. But it is another matter entirely whether we should blame individuals when they display attitudes of disregard toward this moral consideration. On the utilitarian approach, blame should be treated like any other intervention into the causal order, and assessed according to whether it is likely to be optimific in its effects on the welfare of those affected by it. Thus, it is sometimes argued that it would be

wrong to blame affluent individuals when they fail to organize their personal lives according to the principles of utilitarian consumption. Doing so might just discourage and demoralize them, making it clear how onerous morality would be if they actually took it seriously as a basis for ordering their lives, with the effect that those who are blamed would do even less to contribute to improving the conditions of the billions of people whose lives are characterized by deprivation and disease.<sup>17</sup> For these same reasons, it might even be for the best not to encourage children to internalize the kind of emotional commitment to utilitarian standards that would render them disposed to react to deliberate moral infractions with such sentiments as resentment and indignation.<sup>18</sup> The question of the attitudes of an individual toward moral standards thus seems to be completely independent from the question of whether those same attitudes provide others with a normative basis for reactions of blame.<sup>19</sup>

On the internalist approach, by contrast, things are otherwise in this respect. This approach takes the attitudes of agents toward moral considerations to have direct normative significance for the responsibility reactions of others. Moral standards must be ones that we have good reason to become emotionally invested in. There must be something about them that gives an agent's attitudes toward them a special normative significance for our responsibility reactions, protecting from opprobrium those who conscientiously strive to comply with them, and rendering those who are indifferent to them specially vulnerable to blame and opprobrium. Utilitarian and perfectionistic theories seem deficient when viewed in this light. They take moral requirements to derive from an individual's relation to impersonal value or to an ideal of human attainment. But there is nothing in the nature of such requirements that would seem to guarantee that our attitudes toward them have the normative significance for responsibility relations that internalism postulates. Why should other people become exercised by the fact that you are indifferent to an ideal of human perfection, or that you do not take sufficiently seriously the claims made on your behavior by impersonal

value? Theories of this kind might turn out to be correct at the end of the day. At the very least, however, the internalist approach will entail that they are revisionist in some measure, denying an interpersonal dimension of morality that seems to us to be essential to it.

What, by contrast, would a non-revisionist account of moral standards look like? It would be an account of the nature of those standards that makes apparent why peoples' attitudes toward them provide a normative basis for responsibility relations. Most promising in this respect are theories that interpret moral requirements as essentially directional or relational, insofar as they are grounded in the claims or entitlements of other agents. If it is morally impermissible that I do X, on this approach, this is because someone has a claim against me that I not do X. Actions are permissible, by contrast, when nobody has a claim against me that I not perform them. But if it is in the nature of morality that it is relational in this way, then we can immediately see why peoples' attitudes toward moral standards should have direct normative significance for the responsibility reactions. Indifference to moral requirements will in effect be indifference to the claims of other individuals. And we generally have good reason to care about whether people display this attitude in their interactions with each other.

More specifically, we each have compelling reason to care about whether other people are indifferent to moral standards when their attitudes amount to indifference toward *our* moral claims. In these cases, indifference to moral requirements is tantamount to disregard of our own moral standing. The paradigmatic response to such an attitude is resentment, which is an emotion that is called for in cases in which other people act with a lack of consideration for our own claims against them, and which is also the most basic form of moral blame. On the relational conception of morality, attitudes of indifference to moral requirements thus have direct normative significance for the responsibility reactions of those whose claims are disregarded. By the same token, agents achieve normative protection from these basic responsibility reactions

through the conscientious attempt to comply with moral requirements in their dealing with each other. Attitudes of this kind acknowledge the legitimate claims of those affected by the agent's actions, and this in turn undermines the normative basis for resentment.

Other responsibility reactions are, I believe, parasitic on the relational structure that is latent in resentment, and the normative significance of an agent's attitudes for general questions of responsibility operates through their significance for the individuals whose claims are directly affected by them. Thus, indignation and guilt may be understood to be attitudes that are adopted vicariously, on behalf of an individual whose claims have been disregarded by another agent. We feel indignation when we acknowledge that one person has acted with a lack of consideration toward the claims of another, and are subject to guilt in cases in which we acknowledge that we ourselves have done the same. Internalizing a concern for relational requirements is in the first instance a matter of coming to care that our own claims are recognized by others in their interactions with us. But this concern naturally extends to the claims of other agents as well, disposing us to indignation and guilt in cases in which people act with a lack of consideration for those claims. By adopting attitudes of this kind, we therefore open ourselves to the opprobrium not only of those whose claims we disregard, but potentially of the entire moral community as well; conversely, we can protect ourselves from responsibility reactions of this kind through the sincere effort to respect other's claims against us.

Of course, I have done nothing to defend a relational conception of morality in this paper. But it seems to me the kind of moral theory we will be pushed toward if we accept the internalist conception of morality that I have been trying to sketch. And to the extent the internalist approach is found attractive, this will be one important consideration in favor of the relational conception.

---

<sup>1</sup> Compare Stephen L. Darwall, *Impartial Reason* (Ithaca: Cornell University Press, 1983), p. 54.

<sup>2</sup> Moral judgment internalism, as formulated in the text, raises large interpretative questions. To which moral judgments does it apply? How exactly is the notion of motivation to be understood? I shall bracket such issues in my discussion.

<sup>3</sup> For two recent statements, see Simon Blackburn, *Ruling Passions* (Oxford: Clarendon Press, 1998), chap. 3, and Allan Gibbard, *Thinking How to Live* (Cambridge, Mass.: Harvard university Press, 2003), pp. 8-11.

<sup>4</sup> See, for example, David Brink, *Moral Realism and the Foundations of Ethics* (Cambridge, England: Cambridge University Press, 1989), chap. 3; also Sigrun Svavarsdóttir, "Moral Cognitivism and Motivation", *The Philosophical Review* 108 (1999), pp. 161-219.

<sup>5</sup> For this sense, see Scanlon, *What We Owe to Each Other* (Cambridge, Mass.: Harvard University Press, 1998), chap. 1.

<sup>6</sup> R. Jay Wallace, *Responsibility and the Moral Sentiments* (Cambridge, Mass.: Harvard University Press, 1994).

<sup>7</sup> Compare my "Dispassionate Opprobrium. On Blame and the Reactive Sentiments", in R. Jay Wallace, Rahul Kumar, and Samuel Freeman, eds., *Reasons and Recognition. Essays on the Philosophy of T. M. Scanlon* (New York: Oxford University Press, 2011), pp. , sec. 5.

<sup>8</sup> This suggests that it is a kind of conceptual truth that morality has this sort of normative significance for our attitudes of caring and concern; I return to the suggestion at the end of the present section.

<sup>9</sup> Compare the discussion of "no blameworthiness without fault" in my *Responsibility and the Moral Sentiments*, chap. 5.

---

<sup>10</sup> See my *Responsibility and the Moral Sentiments*, chap. 6.

<sup>11</sup> T. M. Scanlon, *Moral Dimensions: Permissibility, Meaning, Blame* (Cambridge, Mass.: Belknap Press of Harvard University Press, 2008), chaps. 1-3.

<sup>12</sup> See Scanlon, *Moral Dimensions*, chap. 4.

<sup>13</sup> For a penetrating discussion of this part of Scanlon's argument, see Niko Kolodny, "Scanlon's Investigation: The Relevance of Intent to Permissibility" REF. The account of the moral excuses I offered in *Responsibility and the Moral Sentiments*, chap. 5 assumed that impermissibility always depends on intent. This now seems to me to be mistaken; the remarks that follow are in part an attempt to explain the significance of the agent's quality of will for questions of blame, given the assumption that impermissibility is at least sometimes independent of the agent's attitudes.

<sup>14</sup> See Scanlon, *What We Owe to Each Other*, chap. 4.

<sup>15</sup> Scanlon, *What We Owe to Each Other*, pp. 18-22.

<sup>16</sup> It is a slightly odd consequence of this account that the consideration that is cited to explain the reason-giving force of morality, namely mutual recognition, is not itself our primary reason for complying with moral standards. Rather, we achieve mutual recognition when we regulate our intentions in ways that acknowledge the prior and independent normative significance of permissibility. This corresponds to the first-person perspective of the agent who manifests attitudes of mutual recognition. Such an agent cares fundamentally about complying with the objective conditions of moral permissibility, not just achieving the subjective conditions for mutual recognition. Thus, if you point out to me that I'm acting in a way that is really objectionable, I won't respond by saying, "That's all right, I'm sincerely trying to do the right thing, and that's all that really matters." Rather, insofar as I'm sincerely trying to do the right thing, I'll be very concerned about whether what I'm doing is morally permissible in fact.

---

<sup>17</sup> See, e.g., Peter Singer, *Practical Ethics*, third edition (New York: Cambridge University Press, 2011), pp. 213-5.

<sup>18</sup> Compare Henry Sidgwick's sympathetic discussion of the possibility of an esoteric morality, in his *The Methods of Ethics*, seventh edition (Indianapolis: Hackett Publishing Company, 1981), book 4, chap. 5.

<sup>19</sup> By the same token, a conscientious effort to comply with the utilitarian principle will notoriously not suffice to provide protection from the opprobrium of others. We can easily imagine circumstances in which it might be optimific, and therefore required by utilitarian lights, to react to such a conscientious agent with the reactions characteristic of blame.