Free Will and Reactive Attitudes
Perspectives on P.F. Strawson’s “Freedom and Resentment”

Edited by
MICHAEL McKENNA
Florida State University, USA

PAUL RUSSELL
University of British Columbia, Canada

ASHGATE
Emotions, Expectations and Responsibility

R. Jay Wallace

In these passages from his book Responsibility and the Moral Sentiments R. Jay Wallace develops a general interpretation of the reactive attitudes, and draws on this interpretation to develop an account of what we are doing when we hold people morally responsible. The interpretation of the reactive sentiments takes as central the phenomena of resentment, guilt, and indignation. These attitudes, Wallace argues, are distinguished by their constitutive connection to the stance of holding someone to an expectation or demand. To adopt this stance toward someone is to be disposed to experience the reactive emotions in case the person should violate the expectation or demand in question, or to believe that their so acting would make it fitting that one experience an emotion from this class. Building on this account, Wallace then argues that to hold someone morally responsible is to hold them to a certain class of expectations—i.e. moral obligations that one accepts—in this way that is distinctively connected to the reactive emotions.

Emotions and Expectations

On P.F. Strawson’s view, emotions such as guilt, resentment, and indignation—what Strawson calls the reactive attitudes—provide the key to understanding moral responsibility and its conditions. I intend to develop this idea by working out an account of the stance of holding someone responsible, in terms of the reactive emotions. Before this can be done, however, it will first be necessary to get clear about the nature of the reactive emotions. What are the essential features of these emotions? What distinguishes them from other forms of moral and non-moral sentiment, and holds them together as a class? Answering these questions will provide the foundation for the account of responsibility and its conditions that follows.

A venerable approach to the reactive emotions holds that they are distinguished by the moral beliefs that give rise to them. Thus Joseph Butler suggested that deliberate resentment is excited by the belief that an injury has been done, where injury is construed as a moral concept, to be distinguished from mere suffering or harm or loss.1 More recently, John Rawls has maintained that the moral sentiments are emotions whose explanation requires the invocation of a moral concept.2 In these

terms, the reactive attitudes of guilt, resentment, and indignation are set apart by the 
kind of moral concept that figures in their explanation; specifically, explanation of 
these reactive emotions must invoke the concept of the right, as distinguished from 
that of the good. The appeal of these suggestions is the connection they postulate 
between the reactive attitudes and a certain kind of belief. The emotions of resentment, 
indignation, and guilt all have a propositional content—one feels indignant about 
something, or guilty for something one has done. It is tempting to characterize 
these contents in moral terms, by holding (for instance) that the reactive emotions 
must be caused by the belief that an injustice or an injury has been done. But the 
temptation should be resisted. It amounts to an excessive moralization of the reactive 
attitudes; for though these attitudes are often—perhaps even standardly—caused by 
distinctively moral beliefs, they do not have to be so caused. It is notoriously the case 
that one can feel guilt, for instance, without sincerely believing oneself to have done 
anything that would amount to a moral injury or an infraction of right.

Reflection on such cases has led some philosophers to sever altogether the 
connection of reactive emotions with beliefs. But if Butler and Rawls go too far in 
the direction of moralizing the reactive emotions, this alternative approach deprives 
us of the resources for acknowledging something that is characteristic of the reactive 
emotions as a class. For by severing their essential connection with beliefs, the 
alternative approach deprives the reactive emotions of their propositional content. 
If it seems plausible that guilt may be experienced in the absence of the belief that 
one is at fault, it seems equally plausible that a state that lacks a propositional object 
would not really be a genuine state of guilt at all. We need a way of explaining the 
distinctive propositional content of the reactive emotions, without characterizing 
that content in exclusively moral terms. This is particularly important if we are to 
 improve on Strawson’s account of responsibility in terms of the reactive emotions: 
a successful interpretation of what it is to hold someone responsible, that draws 
on the reactive emotions, must credit those emotions with propositional objects— 
something Strawson’s own account does not clearly manage to do.

My own account of the reactive emotions aims to solve this problem. My main 
contention is that there is an essential connection between the reactive attitudes and 
a distinctive form of evaluation, or quasi evaluation, that I refer to as holding a 
person to an expectation (or demand). This form of evaluation is not conceptually 
prior to the reactive attitudes, but rather is defined in terms of them: to hold someone 
to an expectation, I maintain, is to be susceptible to the reactive attitudes in one’s 
relations with the person. If this is right, then the task of characterizing the reactive 
attitudes must go hand in hand with the task of characterizing this distinctive form of 
 quasi evaluation, for the two can only be understood in terms of each other. Among 
other things, pursuing the connection between reactive attitudes and expectations 
should help us to account for the characteristic propositional content of the reactive 
attitudes, without falling into the trap of overmoralizing them from the start. To 
be in a state of reactive emotion, one must believe that a person has violated some 
expectation that one holds the person to; and in terms of this belief, we can give an 
account of how the reactive emotions have the kind of propositional content that 
distinguishes them from other emotional states. But it need not be the case that the 
expectation that gives the content of a reactive emotion is a moral one, or even that
it is an expectation one sincerely endorses. (This is why I refer to the attitude of holding a person to an expectation as a form of quasi evaluation.)

The Approach Sketched

To start with, note that there is one way of expecting a thing to happen that does not have any special connection with morality, or with the moral emotions. This is the sense in which to expect something to happen is simply to believe there is a high probability that the expected event will occur. Thus when I hit the appropriate button on the remote control, I expect my television to turn on; as the summer draws to a close, I expect that classes will soon begin; and when I assign the Prolegomena to my beginning students, I expect that they will not understand it on their own. Expectations in this sense are often associated with emotions of various kinds. For example, my expectation about the start of classes may be suffused with a feeling of anxiety that has its roots in my childhood experiences of school; the failure of my TV to go on as expected when I activate the remote control may provoke a fit of rage and frustration. But it is not in general the case that expectations of this sort—that is, beliefs about the future—are presumptively associated with any particular attitude. I may equally contemplate the expected start of classes with depression, enthusiasm, or with complete indifference, and none of these emotional responses would necessarily be more fitting than the others.

There is, however, a different way of expecting something to occur that is essentially tied to particular emotional responses. This is the sense in which, as we might say, we "hold someone to an expectation", or in which we demand of people that they act as we expect them to. In the case of my students, for instance, I not only expect, in the first sense, that they will not understand the Prolegomena, I also hold them to the expectation that they will not lie, cheat, attempt to blackmail me or their fellow students, and so on. In holding them to these various expectations, I often believe that the expectations will be fulfilled. Thus I generally believe that my students will not in fact attempt to blackmail me. But even when a belief of this sort is present, it does not capture what is centrally involved in holding a person to an expectation, or in making a demand of the person. The crucial element, I would suggest, is attitudinal: to hold someone to an expectation is essentially to be susceptible a certain range of emotions in the case that the expectation is not fulfilled, or to believe that the violation of the expectation would make it appropriate for one to be subject to those emotions. For reasons that will become clear in the course of my discussion, we may refer to this stance of holding someone to an expectation as a "quasi-evaluative" stance.

Emotions that are constitutively linked to expectations, in this sense of holding someone to an expectation, are the reactive attitudes, as I will interpret them. Take the central examples of resentment, moral indignation, and guilt. These are not mere feelings that one might happen to be subject to in any circumstances whatsoever. I may dislike my television set or be frustrated and annoyed when it fails to turn on; but insofar as I do not hold the TV to expectations, I cannot, properly speaking, be said to resent it or to be indignant at it. Resentment, indignation, and guilt are
essentially tied to expectations that we hold ourselves and others to; susceptibility to these emotions is what constitutes holding someone to an expectation. This mutual dependence of emotion and expectation distinguishes the reactive attitudes, on the account of which I defend—a class of attitudes that includes a central group of moral emotions, but that also includes some nonmoral emotions as well.

Before we can develop this suggestion, however, a word or two of clarification is in order. So far I have used the terms “expectation” and “demand” in sketching the stance that is constitutively connected with the reactive attitudes. I have chosen these terms because they point at once in two different directions, in a way that is appropriate to the psychological phenomena I wish to call attention to.\(^3\) Thus a demand is both a psychological stance we might adopt toward someone and that which we demand of the person when we adopt the stance. Strictly speaking, however, when I write of “holding someone to a demand or expectation”, “demand” and “expectation” are to be taken in the latter sense: that which we demand or expect of the person. They are thus equivalent ways of expressing the notion of a practical requirement or prohibition in a particular situation of action. So construed, expectations must in principle be capable of being formulated linguistically, though for my purposes it will not be important to settle on a canonical formula for expressing such expectations in words. We might express an expectation by using the concepts of prohibition or requirement explicitly, say as an operator on sentences that describe kinds of action in a particular situation (for instance: “It is prohibited that you should break the promise you made to your sister”). Alternatively, expectations might be expressed by imperatives (“Do not break the promise you made to your sister”).

However we choose to express expectations, when so construed there are several further points about them that must be noted. First, it is important that expectations should be capable of being supported by practical reasons. It is not necessary that expectations, to count as such, should either be or be believed to be supported by such reasons; as I will show, it is an important fact about us that we often hold ourselves and others to expectations that are supported by no justification that we may accept. But expectations can be supported by justifications that the agent accepts, and in the favorable cases they are supported in this way, so the possibility of such support needs to be left open. Second, though expectations must be capable of linguistic formulation, it is not required that the agents who hold themselves and others to such expectations should always be in a privileged position to produce such a formulation. Thus an agent might not be aware that she holds other people to a given demand, and might only come to grasp that demand discursively by inference from patterns in her emotional life (involving, for instance, the kinds of situations that move her to resentment). Finally, we should allow the possibility that expectations may conflict, so that in a given situation there are mutually inconsistent kinds of action that one is required to perform. At least, nothing in the very concept of an expectation, as I construe it, should rule out the possibility of such conflicts.

Something must now be said about the connection between expectations, construed in this way, and the reactive emotions. In characterizing this connection, I have used a disjunctive formulation: to hold someone to an expectation, I suggested,
is to be susceptible to a certain level of emotions if the expectation is violated or to believe that it would be appropriate for one to feel those emotions if the expectation is violated. This disjunction is to be understood nonexclusively. Thus there are three different ways in which one might count as having the attitude of quasi evaluation that I have referred to as “holding someone to an expectation”: (1) one simply finds oneself reacting with the emotions of resentment, indignation, or guilt in respect to the violation of a certain set of expectations, and it is part of the explanation of the emotions to which one is thus subject that one believes the expectations in question have been breached; (2) one is not consistently subject to these various emotions oneself on occasions when a given set of expectations has been violated, but one believes that it would be appropriate for one to react to their violation with this range of emotions, and one thinks that what would make these emotional reactions appropriate is the fact that the expectations have been breached; (3) one feels the emotions of resentment, indignation, and guilt on occasions when the expectations have been violated and one believes that these are appropriate emotions for one to feel on such occasions, because the expectations have been breached.

Holding someone to a demand, in any of these ways, can be either a long-term or a short-term condition. One might hold a person to a given demand only long enough to feel momentary resentment toward the person, or one might hold the person to a range of demands over a period of many years. Notice, too, that beliefs figure prominently and variously in my explanation of the stance of holding someone to a demand. There is, first, the belief that someone’s violation of a demand would make one of the reactive emotions appropriate. Having such a belief is not necessary to count as holding the person to the demand (compare case 1), but even in the absence of the appropriate reactive emotion, this belief would be sufficient to qualify one as holding the person to the demand (case 2).

Moreover, when one is subject to a state of reactive emotion, the preceding account entails that a different kind of belief will figure in the explanation of the emotion. This is the belief that some demand has been breached. Even if we choose to express demands as imperatives, and hence as lacking truth-conditions, there will still be room for the belief that a demand has been violated. Furthermore, some belief of this kind must be present whenever an agent is in a particular state of reactive emotion, and must contribute to explaining why one is in this state.

This explanatory role of beliefs in accounting for particular states of reactive emotion is extremely important, since it will eventually provide the key to understanding the propositional content characteristic of the reactive emotions as a class. For the present, though, it will suffice to make the following observations. The stance of holding someone to a demand, as I interpret it, does not have explanatory priority

---

4 By “belief” here, I mean an exclusively cognitive state whose propositional content is directly assessable as true or false. In these terms, a state in which one entertains a proposition one does not fully accept—to which one assigns an extremely low probability, for instance—may be allowed to count as a belief, at least in a degenerate sense. For the suggestion that the content of many emotions can best be understood in terms of beliefs of this degenerate sort, see P.S. Greenspan, *Emotions and Reasons: An Inquiry into Emotional Justification* (New York: Routledge, Chapman and Hall, 1988).
vis-à-vis the reactive emotions: to be subject to the reactive emotions is to take this stance toward a person, and to adopt this stance is in turn to be subject to the reactive emotions. At the same time, the nature of the connection between the reactive emotions and this stance is such that particular states of reactive emotion must always be explicable in terms of some belief concerning the violation of a demand. In this way, beliefs about the violation of demands have a kind of priority in accounting for particular states of reactive emotion.

Regarding particular states of reactive emotion, it should be further noted that I have not provided, and do not intend to provide, a complete analysis of what it is to be in such an emotional state. For purposes of discussion, we may assume that each of the reactive emotions is associated with a distinctive syndrome of sensory, behavioral, and linguistic dispositions. The state of guilt, for instance, is associated with characteristic patterns of salience and attention, dispositions to action, expectations regarding the reactions of others, susceptibilities to feel certain sorts of sensation, and so on. Such syndromes may be taken to provide a rough characterization of the various emotional states. But it is not clear that these characterizations can be regarded as complete analyses of the various emotional states, in the sense of providing a specification of sufficient conditions for being in those states. For my purposes, it is not important that we have an analysis in this sense available.

What I do wish to insist on is that there is one characteristic that is essential to the reactive emotions, and that may be taken to distinguish them as a class from other types of attitudes. This is their connection to expectations, the connection I have described in terms of the quasi-evaluative stance of holding someone to an expectation. To pursue this strategy for understanding the reactive attitude is to define them together with the stance of quasi evaluation. The quasi-evaluative stance of holding someone to an expectation is characterized in terms of a susceptibility to the reactive emotions. And those emotions in their turn are characterized in terms of this evaluative stance, as the emotions one is susceptible to in virtue of holding someone to an expectation. Since neither of these items is taken to be conceptually prior, the strategy will only be illuminating if they can both be characterized together in ways that differentiate them from other evaluative stances, and that help us to mark important and recognizable distinctions between the reactive attitudes and other kinds of (moral and nonmoral) emotions.

Reactive and Nonreactive, Moral and Nonmoral

The reactive attitudes of resentment, indignation, and guilt all seem to have fairly complex propositional objects: one feels resentful or indignant about something that somebody has done, or guilty for having done something oneself. To account adequately for this aspect of the reactive emotions, we must suppose them to have a cognitive dimension; in particular, it seems that a person subject to a reactive emotion must have some kind of evaluative belief, one that figures in the explanation

---

5 See, for example, Rawls, *A Theory of Justice*, sec. 73.
of the emotional state. Reactive attitudes are also often moral sentiments, in that the evaluative beliefs that give rise to them are often beliefs that some moral transgression has been committed (by oneself or some other party), and yet they are not the only moral sentiments to which people are subject, nor are they exclusively moral sentiments. We need an account of the cognitive dimension in reactive attitudes that will enable us to draw the right kind of line: between the moral and the nonmoral reactive attitudes, and between moral reactive attitudes and other kinds of moral sentiment.

The approach I have sketched suggests a straightforward characterization of the cognitive dimension of reactive attitudes. On that approach, reactive attitudes as a class are distinguished by their connection with expectations, so that any particular state of reactive emotion must be explained by the belief that some expectation has been breached. It is the explanatory role of such beliefs about the violation of an expectation that is the defining characteristic of the states of reactive emotion as a class, and that provides them with their distinctive propositional objects; beliefs of this sort will therefore always be present when one is in one of the reactive states. Take indignation: a particular state of indignation will be focused on a specific propositional object; there must be something about which one is indignant, if the emotional state one is in is to count as indignation at all. According to the approach I have sketched, this propositional object of indignation can be specified by the belief that some expectation one holds people to has been breached. And what entitles us to suppose that this belief specifies the propositional content of the emotional state is the fact that it explains why one is in the emotional state.

Of course, it is possible to have the belief that an expectation has been violated without being in one of the reactive states, and so such beliefs will not always be sufficient conditions, by themselves, to produce a state of reactive emotion. One might believe that a demand has been violated that one does not hold people to oneself, in the sense I have been trying to characterize, and in this case the belief that the demand has been violated would not give rise to one of the reactive emotional states. How then, to characterize the difference between the agent for whom beliefs about the violation of prohibitions or requirements suffice to produce states of reactive emotion, and the agent for whom they do not? It is tempting to say that the difference consists in the stance of the two agents toward the expectations in question: the first agent adopts a stance toward the expectations, in virtue of which we can say that she holds people to them in a way that the second person does not. This is so far correct, but it should be borne in mind that the relevant stance is not necessarily something distinct from a susceptibility to states of reactive emotion when the expectations in question are breached. For this reason we should not say that the stance together with an agent’s belief that an expectation has been breached—explains the agent’s state of reactive emotion. Rather, that the agent has the stance displays itself in the fact that beliefs about the violation of certain expectations give rise to states of reactive emotion.

If this account of the cognitive element in reactive emotions is right, it suggests a natural way of distinguishing between moral and nonmoral reactive attitudes as well as a way of distinguishing the moral reactive attitudes from other kinds of moral emotions. I will start with the distinction within the category of reactive attitudes,
between the moral and the nonmoral cases. Whether a reactive emotion is or is not a moral one would seem to depend on the kind of belief that gives rise to it, and hence the kind of propositional content that it has. At least on the face of it, there is no reason to suppose that all of the expectations we hold people to are distinctively moral in character, and so we might distinguish between the belief that a moral expectation has been violated and the belief that some nonmoral expectation has been violated. In these terms, a moral reactive emotion (say, resentment) would be one that is explained by a belief of the first type and therefore has a distinctively moral content, while nonmoral resentment would have a content specified by beliefs of the second (nonmoral) type.

This strategy for distinguishing between moral and nonmoral reactive attitudes may usefully be compared with Strawson’s way of carving up the reactive attitudes, which also differentiates between moral and nonmoral varieties. Strawson identifies three different kinds of reactive attitudes. What he calls personal reactive attitudes are “the non-detached attitudes and reactions of people directly involved in transactions with each other”, where one demands of others a degree of consideration toward oneself; resentment is the primary example here, but Strawson also mentions gratitude, love, forgiveness, and hurt feelings. The distinctively moral reactive attitudes, Strawson seems to think, are not these personal attitudes, but vicarious analogues of them, such as moral indignation or disapprobation, where one holds people to standards of behavior and attitude, not specifically in relation to oneself, but in regard to others. Finally there are self-reactive attitudes, associated with demands made on oneself in one’s conduct with others; examples include guilt, remorse, shame, and the sense of obligation.

Taken literally, however—and ignoring for the moment Strawson’s inclusive interpretation of reactivity—this approach has some peculiar consequences. It would rule out the possibility of agent’s feeling moral resentment on her own behalf, about the violation of moral expectations by other people in their behavior toward the agent; resentment about being treated unfairly, for instance, would not count as a moral sentiment. More strangely still, Strawson’s approach classifies guilt as an exclusively nonmoral sentiment, since it involves the imposition of demands


7 FR, 28–9/70–71. Strictly speaking, Strawson says only that the moral reactive emotions must be “capable” of being vicarious; but he does not explain how a nonvicarious emotion might be potentially vicarious. My alternative account of moral reactive emotions suggests a natural unpacking of this idea: a personal emotion may be capable of being vicarious if the demand it rests on is supported by reasons that generalize to cases that do not directly involve the agent subject to the emotion.

8 FR, 28–30/71–2.
on oneself rather than on other parties.⁹ We get a more recognizable division of the reactive attitudes into moral and nonmoral varieties if we follow the suggestion I have made, that what makes a reactive attitude a moral one is not its vicarious quality but the kind of expectation it is essentially bound up with. The most plausible development of this suggestion would begin by noting that many expectations we hold people to are supported by justifications, which identify reasons for complying with the expectations. Moral expectations can then be defined as expectations that are justifiable in terms of distinctively moral reasons.¹⁰ I will refer to expectations that admit of this kind of moral justification as obligations.

I propose that reactive emotions be classified as moral when they are connected with moral obligations in this sense. More precisely, we should count reactive emotions as moral when they are linked to obligations for which the agent is herself able to provide moral justifications; these justifications identify reasons that explain the agent’s own efforts to comply with the obligations in question, and they provide moral terms that the agent is prepared to use to justify such compliance on the part of others, whom the agent holds to the obligations.¹¹ When they are linked with obligations of this kind, it is natural to treat reactive emotions as moral sentiments, since their explanation essentially requires moral beliefs, namely beliefs about the violation of what the agent himself correctly regards as moral obligations. The explanatory role of such moral beliefs gives these emotional states a distinctively moral content. And in fact we commonly do regard resentment, indignation, and guilt to be moral emotions when they are incited by beliefs about the violation of moral obligations.¹²

It is a further consequence of this general approach, however, that not all reactive attitudes need be distinctively moral in this way. Insofar as some of the expectations to which we hold ourselves and others are not moral obligations (that is, expectations supported by moral justifications), the reactive emotions to which they give rise will

---

⁹ A similar point is made by Jonathan Bennett, who suggests that the two basic categories of reactive attitudes are the “principle” and the “nonprinciple” (“Accountability”, in Philosophical Subjects, ed. Zak van Straaten [Oxford: Oxford University Press, 1980], pp. 45–7). ... In his “Reply to Ayer and Bennett”, Strawson admits that it was an error “so to use the word ‘moral’ as to exclude self-reactive attitudes from its scope” (in van Straaten (ed.), p. 266). But he, like Bennett, does not seem to acknowledge that personal reactive attitudes might also be moral sentiments, depending on the character of the expectations with which they are connected.

¹⁰ It is not necessary to opt here for a particular account of what makes reasons moral. For purposes of discussion in this book, however, I shall assume that moral reasons need not be impersonal or “agent-neutral” but may also be “agent-relative” (containing an essential reference to the particular agent who has the reasons). On the distinction between agent-neutral and agent-relative reasons, and the implications of the distinction for issues of objectivity in ethics, see Thomas Nagel, The View from Nowhere (New York: Oxford University Press, 1986), Chs 8–9.

¹¹ The moral obligations in question would thus be ones that the agent accepts ....

¹² Many of the demands we hold people to are supported by both moral and nonmoral reasons. In these cases we might say that the emotions caused by the violation of the demands are moral insofar as the reasons that support those demands are moral.
not have a distinctively moral content. There is nothing peculiar in the supposition that we might resent a breach of etiquette in the behavior of another person toward us, or that we might feel guilty about our failure to measure up to expectations that we have taken over from our parents or our church but that we do not consider to be supported by any justifications at all. In both cases, however, it is required that there be some belief about the breach of an expectation that figures in the explanation of the reactive emotion. In this way we can begin to move away from the excessive moralization of the reactive attitudes, without denying them the kind of propositional content that sets them apart from other emotional states.

It remains to say something about the distinction between moral reactive attitudes and other types of moral sentiment. Here again it would seem that the beliefs that give rise to the emotions of various kinds may provide the key. Thus moral reactive attitudes are explained by the belief that some moral obligation has been violated. But this is not the only kind of moral belief that we entertain. Consider the various modalities of moral value, such as the values of kindness or consideration or benevolence or even justice. Such values often coincide with our moral obligations. Just as we demand, for example, that people keep their promises, or help others in extreme distress, so may we value acts of fidelity and benevolence, regarding them as good and admirable. But moral values can also diverge from moral obligations, construed as strict prohibitions and requirements. For instance, we may think that a certain sort of character is especially virtuous, even though we do not, strictly speaking, demand of people that they exhibit a character of that sort; or we might think that a particular action displays a degree of beneficence or consideration that goes well beyond what we actually demand of each other in our normal interactions. In all of these cases—the cases in which moral values coincide with our moral obligations, and the cases in which they exceed those obligations—there are often characteristic moral sentiments that are caused by evaluative moral beliefs. In addition to feeling guilt about my failure to act in accordance with the demands to which I hold myself, I may feel moral shame because I lack the moral excellences that I aspire to. And I may feel gratitude toward someone whose actions toward me are unusually beneficent or admire someone whose character is virtuous to an exemplary degree. Thus we may distinguish moral reactive emotions from other moral sentiments in terms of the kinds of moral beliefs that give rise to the moral sentiments, and that fix the content of those sentiments: the reactive attitudes are explained exclusively by beliefs about the violation of moral obligations (construed as strict prohibitions or requirements), whereas other moral sentiments are explained by beliefs about the various modalities of moral value.14

13 Some such breaches of etiquette will be resented, because they violate a distinctively moral obligation of respect or consideration. But I am imagining a case in which resentment is occasioned solely by the belief that a requirement of etiquette has been violated, not by a distinct moral belief.

14 Compare Rawls, A Theory of Justice, p. 484: "In general, guilt, resentment, and indignation invoke the concept of right, whereas shame, contempt, and derision appeal to the concept of goodness."
In the previous section I suggested that it is at least conceivable that there might be cultures whose members do not have the stance of holding people to expectations in their repertoire. The distinction just drawn between moral reactive attitudes and other kinds of moral emotions may help to flesh out this suggestion. The nonreactive moral emotions, I have suggested, are connected with beliefs about the various modalities of moral value. A shared set of such values might conceivably have sufficient structure to constitute an ethical system for regulating social interactions within a culture, and yet the members of that culture might not be subject to the distinctively reactive emotions at all. Perhaps they respond to their own failure to live up to their values with shame rather than guilt, and hold others who similarly fall short of such standards in contempt or derision, rather than resenting them or feeling indignation; a susceptibility to these nonreactive emotions might be sufficient to guide their conduct and to provide a framework for some kind of common ethical discourse.

It has frequently been claimed that there are "shame cultures" of this kind, whose members lack the characteristic emotional resources of guilt and the other reactive attitudes. Whether such claims are true is not a question that needs to be decided here. For my purposes it is sufficient to note the following points. First, there seems to be nothing in the very idea of the reactive emotions that rules out the possibility of shame cultures. The distinctive features of resentment, indignation, and guilt do not seem to be given along with the bare facts of human social life, and so one can see how there might be human communities whose members are not subject to these emotions. Second, a shame culture would not necessarily be one in which there are no recognizable ethical norms, construed as norms that make social cooperation possible; nor would it be the case that the members of such a culture would not internalize these norms, in the sense of having incentives for compliance with the norms whose effectiveness is potentially independent of externally administered sanctions and rewards. On the contrary, the plausibility of the claim that there are shame cultures seems to me to rest crucially on the assumption that conformity with ethical norms can be sustained by emotional resources that do not include the

15 The claim that there are, in fact, cultures without the reactive emotions presupposes that such emotions cannot be understood in exclusively biological terms, but that they are instead somehow culturally constituted. For a summary discussion of endogenous (biological) and exogenous (cultural) accounts of emotions, see Paul Heelas, "Emotions across Cultures: Objectivity and Cultural Divergence", in S.C. Brown (ed.), Objectivity and Cultural Divergence, Philosophy (supp.) 17 (1984), pp. 21-42. The idea that reactive emotions are culturally constituted presupposes that susceptibility to say, guilt requires the ability to make attributions of guilt; but it does not presuppose the stronger and less plausible claim that one can only be in a state of guilt if one believes of oneself that one is in such a state at the time. Allan Gibbard seems to slide between these two claims when discussing what he calls "attributional" accounts of emotion in his book Wise Choices, Apt Feelings: A Theory of Normative Judgment (Cambridge, MA: Harvard University Press, 1990), pp. 141-50. (The less plausible claim has been advocated by Stanley Schachter, in Emotion, Obesity, and Crime (New York: Academic Press, 1971). For a criticism of Schachter's interpretation of his experimental results, see Robert M. Gordon, The Structure of Emotions: Investigitations in Cognitive Philosophy (Cambridge: Cambridge University Press, 1987), Ch. 5.)
distinctively reactive emotions—that conformity can be sustained by the internal sanctions of shame and anger, for instance, rather than those of guilt and indignation.\textsuperscript{16} This assumption may be empirically ungrounded, but there seems to be no reason to reject it on conceptual grounds alone. This is what I meant in saying that nothing in the very idea of the reactive emotions rules out the possibility of shame cultures.

But even if it should turn out that there are in fact no pure shame cultures, we should still distinguish the different ways in which emotions contribute to sustaining social cooperation. What I have called the reactive emotions differ from such emotions as shame and anger in their presumptive connection with the kind of prohibitions or requirements that I have referred to as moral obligations. These emotions help to define what Bernard Williams has called “the morality system”, to mark a contrast between a conception of the moral and its demands that is especially prominent in modern, Christianized cultures and other aspects or forms of ethical life.\textsuperscript{17} Though I do not share Williams’s evident hostility toward the morality system so construed, I agree that it represents a distinctive interpretation of ethical prohibitions and requirements, and in the next chapter I shall argue that the special connection it postulates between moral obligations and the reactive emotions is the key to understanding what it is to hold someone morally responsible.

Responsibility

The question of what we are doing when we hold people responsible has not been adequately treated in discussions of freedom and responsibility. A common assumption is that moral responsibility can be understood primarily in terms of moral blame and sanction, so that to hold people morally responsible is to be prepared to blame or sanction them for their moral offenses, where the sanctions tend, at the limit, toward punishment. This is all right, so far as it goes, but philosophers have not yet given us a satisfactory interpretation of the stance that issues in these forms of treatment.

I take up this problem below, developing an approach to moral responsibility that builds on the account of the reactive emotions defended in the preceding chapter. According to that account, the reactive emotions of resentment, indignation, and guilt are distinguished by their connection with expectations (construed as prohibitions or requirements); so that to hold someone to such an expectation is to be susceptible to

\textsuperscript{16} The distinction between shame and guilt cultures does not necessarily carry with it the implication that shame cultures are psychologically and morally primitive, by comparison with guilt cultures. Indeed, shame-based moralities have recently been defended as superior to systems centered around the reactive emotions: see Bernard Williams, \textit{Shame and Necessity} (Berkeley: University of California Press, 1993), Ch. 4; and, in a similar vein, Annette C. Baier, “Moralism and Cruelty: Reflections on Hume and Kant”, \textit{Ethics} 103 (1993), pp. 436–57. I take issue with some of Williams’s and Baier’s criticisms of guilt-based moralities in the chapters to follow; but they are surely correct to challenge the complacent “progressivism” that characterized many earlier discussions of the distinction between guilt and shame cultures.

\textsuperscript{17} See Bernard Williams, \textit{Ethics and the Limits of Philosophy} (Cambridge, MA: Harvard University Press, 1985), Ch. 10.
the reactive emotions in the case that the expectation is breached, or to believe that the reactive emotions would be appropriate in that case. To hold a person morally responsible, I now want to argue, is to hold the person to moral expectations that one accepts. This approach correctly treats holding people responsible as a basic stance that we take toward them, that cannot be reduced to a behavioral disposition to sanction them for what they do. Furthermore, it connects the practice of holding people morally responsible with the notions of moral obligation, moral right, and moral wrong, which form the nexus of concepts in which responsibility would seem to belong. At the same time, the interpretation I offer suggests an appealing explanation of how moral blame and the various sanctioning responses hang together as a class: to blame someone is to be subject to one of the reactive emotions in terms of which the stance of holding people responsible is essentially defined, and these emotions are expressed by the sanctioning behavior to which the stance of holding people responsible inclines us. This point is developed below in the section, “The Reactive Account and Moral Judgment”, which considers the question of how judgments of moral responsibility may be understood to go beyond mere descriptions of what an agent has done, and explains how we may hold someone responsible for a moral wrong without actually being subject to an episode of reactive emotion.

Responsibility and the Reactive Emotions

Above I argued that the reactive emotions of resentment, guilt, and indignation should be understood in terms of the quasi-evaluative stance of holding people to expectations. I now want to suggest that this stance provides the key to understanding what we are doing when we hold people morally responsible.

Holding someone to an expectation has been characterized in terms of the reactive emotions: to hold someone to an expectation is to be susceptible to the reactive emotions, or to believe that it would be appropriate for one to feel the reactive emotions, in the case that the expectation is violated. Now it would seem that when we hold people morally responsible we are similarly susceptible to the reactive emotions, if those held responsible breach our expectations, that we believe it would be appropriate for us to feel the reactive emotions in those cases. This suggests that moral responsibility might be analyzed in terms of the quasi-evaluative stance of holding people to expectations. To hold a person responsible, we might suppose, is simply to hold the person to expectations in the way that is connected with the reactive emotions.

This is a promising suggestion. The reactive emotions of indignation, resentment, and guilt seem to be natural candidates for the attitudinal component of moral blame, and with this component in place, one could understand the various moral sanctions to be unified by their common function of expressing the reactive emotions. But as formulated so far, this account cannot be correct, for two reasons. First when we hold people morally responsible, we are interested in whether their behavior does or does not comply with distinctively moral requirements. But it is possible to hold people to expectations that are not supported by specially moral justifications; in these cases, then, the stance of holding someone to an expectation would seem to range more
widely than the stance of holding someone morally responsible. A second and more important possibility of divergence between the two stances is presented by cases of irrational guilt and resentment. As explained in Chapter 2, such cases should be understood as cases where one holds oneself or others to expectations that one does not fully accept, for purposes of practical deliberation and normative discussion. But if we do not accept a given set of expectations, I do not think that we would hold ourselves or others morally responsible for the failure to comply with them. A person who feels irrational guilt, for instance about the violation of a parental prohibition on going to the movies, is not apt to blame herself for the action that prompts the guilt, but to view the guilt as a symptom to be treated and cured. Moral responsibility seems to be tied to distinctively moral expectations, which are supported by reasons that we ourselves accept as a basis for practical deliberation and normative criticism and discussion.

This suggests the following revision of the initial account: to hold someone morally responsible is to hold the person to moral expectations that one accepts. The set of moral expectations that one both accepts and holds people to is basically the class of what I earlier called moral obligations. Restricting the set of expectations to moral obligations that the agent accepts focuses the analysis correctly on the range of cases in which moral responsibility would seem to come into play. To see this, recall that the notion of an expectation, as introduced in Chapter 2 (not included in this collection), was meant to capture the idea of a prohibition or requirement. Hence the moral obligations that one accepts, and that one holds oneself and others to, mark out a class of distinctively moral prohibitions or requirements. This class of moral prohibitions and requirements that one both accepts and holds people to constitutes a special sphere within our ethical concepts. It is the sphere of moral rightness or wrongness—a sphere that is narrower than the sphere of morality as a whole, but broader than the sphere of justice and injustice. The sphere is narrower than the sphere of morality as a whole, because there are moral considerations that confer value on actions and character traits without being strictly matters of moral obligation, right, or wrong. For instance, a person who is superlatively beneficent will act in ways that go beyond our moral expectations but that are still morally valuable; this is the sphere of the supererogatory. At the same time, the sphere of moral right and wrong encompasses more than considerations of justice and injustice alone, since there are things that it would be morally wrong to do, such as causing others unnecessary suffering, that may not strictly violate requirements of justice. The term "obligation", as I have introduced it, captures the set of moral requirements of right that an agent accepts.\(^{18}\)

\(^{18}\) In using "obligation" to refer to all the requirements of right, I depart from some other conventions that have been adopted. Rawls, for instance, restricts the term to requirements on individuals that derive from voluntary acts whereby those individuals benefit from a just institution or practice; see John Rawls, *A Theory of Justice* (Cambridge, MA: Harvard University Press, 1971), sec. 18. By contrast, "obligation" as I shall use it ranges over all the requirements of right on individuals, including what Rawls calls natural duties, as well as obligations in his sense.
The account I am developing thus situates our practice of holding people morally responsible within a distinctive nexus of moral concepts, namely those of moral obligation, moral right, and moral wrong. This seems to be the context in which moral responsibility properly belongs, for judgments of responsibility, with their characteristic connection to moral blame and moral sanction, would appear to come into play primarily in cases where people have violated the moral obligations we hold them to. The special connection between moral obligation and moral responsibility and blame has been remarked by G. E. M. Anscombe and Bernard Williams, both of whom trace the strict notion of moral obligation to theological ideas that no longer have a secure place within contemporary moral life.\(^{19}\) Anscombe contends that the notion of moral obligation is only intelligible within the context of a divine law conception of ethics, which we do not now accept, and Williams suggests that the notion of obligation is the central normative concept of “the morality system”, which he takes to be a simplifying and aggressive interpretation of our ethical ideas under pressure of recognizably Christian concerns.

But even if Anscombe and Williams are correct in thinking that the notion of moral obligation was originally part of a distinctively theological conception of ethics, it does not follow that it is only intelligible within the context of religious ideas. Whether it is or is not depends on whether we can find an interpretation of moral obligation, right, and wrong and, I should add, the associated ideas of moral responsibility, blame, and sanction that relates them to a secular understanding of human nature and practical reason (a project begun, though hardly completed, by the moral philosophers of the modern period, from Hobbes through Kant). So the genealogical association of moral responsibility with theological assumptions does not necessarily call into question the modern practice of moral responsibility. If the genealogical proposal is correct, however, it implies that the members of societies that never had the relevant theological ideas could not, strictly speaking, be said to hold each other morally responsible. This seems peculiar, since it is natural to think of such people (the Athenian contemporaries of Plato and Aristotle, for example) as having had available to them the responses of blame and moral sanction.\(^{20}\) What are we to make of this thought?

The genealogical suggestion seems to me to go together with the idea, broached in the preceding chapter, that there might be shame cultures whose members do not have the reactive emotions in their repertoire. The moral reactive emotions will only be available where the moral notions of obligation, right, and wrong are in place, and if Anscombe and Williams are correct, those ideas are historically and culturally local, linked to theological outlooks that are far from universal. But to suppose that there are cultures without the reactive emotions is not to suppose that the members of


20 Aristotle’s famous discussion of the voluntary, for instance, seems to presuppose a form of moral assessment that is “deep”, in that it goes beyond mere evaluative description of what an agent has done; see The Nicomachean Ethics, bk. 3, Chs 1 and 5.
those cultures are altogether lacking in moral emotions. On the contrary, the idea of cultures without the reactive emotions seems plausible only on the assumption that other moral emotions are available in such cultures, such as shame and anger, and that these emotions are capable of providing internal sanctions sufficient to motivate general compliance with a system of ethical norms. But if the members of a shame culture are susceptible to motivating emotions of this sort, we can understand how something analogous to our practice of moral responsibility might emerge in such a culture. The analogous practice, like ours, would involve the responses of blame and moral sanction, but these responses would be understood in terms of the different moral emotions that facilitate social cooperation within the shame culture: blame would involve the emotions of shame and anger, rather than the reactive emotions of resentment, indignation, and guilt, and moral sanctions would serve to express these nonreactive emotions.  

This approach postulates a generic stance of holding people morally responsible that is defined in terms of whatever moral sentiments prevail within a given culture, and so is available even where the theological ideas identified by Anscombe and Williams may not have left their traces. Our practice of holding people morally responsible, centered as it is around the notions of moral obligation, right, and wrong, would then be distinguished by its connection to a specific subset of moral sentiments, namely the reactive emotions of resentment, indignation, and guilt.  

In this way, we preserve the privileged connection between our practice of holding people responsible and the notions of obligation, right, and wrong, while allowing that cultures without these moral ideas could have analogous patterns of response to moral offenses.

What is immediately striking about this approach is the role it ascribes to the moral emotions. The approach says that we hold people morally responsible for complying with moral obligations only if we hold them to those obligations. Since holding people to an expectation is in turn understood in terms its connection with the reactive emotions, the strategy essentially links moral responsibility to the moral reactive emotions. Specifically, one may say that the basic stance of holding someone morally responsible involves a susceptibility to reactive emotions if the person breaches moral obligations that we accept, or the belief that it would be appropriate for us to feel those emotions if the person should violate those obligations.  

---

21 Bernard Williams has defended the idea that the ancient Greeks had a sophisticated conception of moral responsibility—analogue to distinctively modern conceptions, and in many ways superior to such conceptions—even without such reactive emotions as guilt; see his Shame and Necessity (Berkeley: University of California Press, 1993), especially Ch. 3.

22 I should emphasize that I have not taken a stand on the genealogical hypotheses of Anscombe and Williams, or on the related idea that there are shame cultures that lack the reactive emotions of resentment, indignation, and guilt. My aim has rather been to explore the implications of these proposals for the account of moral responsibility I have been developing. If the proposals should turn out to be incorrect, then what I have called the "generic" stance of holding people morally responsible would include only the familiar modern form of responsibility, connected with the reactive emotions.

23 Of course, which reactive emotions would be appropriate depends on who has violated the obligations, and how the act that violates the obligations is related to the subject of the emotions. Guilt is appropriate to one's own violations, indignation to violations by others, and
of the central role this account ascribes to the reactive emotions, I will refer to it as the reactive account of moral responsibility.

By connecting moral responsibility to the reactive emotions in this way, the reactive account promises to improve on the economy of threat approach .... It treats the stance of holding people responsible essentially in terms of attitudinal conditions, and so avoids the behavioristic danger of associating responsibility too exclusively with moral sanctions. To hold myself responsible for a moral wrong for example, it is sufficient that I should feel guilt about my violation of a moral obligation that I accept, or at least believe that that violation would make it fitting for me to feel guilt; these conditions could of course be satisfied without my expressing my guilt to myself or anyone else, and certainly without my sanctioning myself. Furthermore, the reactive emotions seem to have the right kind of content to capture the attitudinal dimension of moral blame. On the account of them I have offered, resentment, indignation, and guilt are backward-looking emotions, responses to the actions of a particular agent (or agents); they are essentially about such actions, in a way that exactly captures the backward-looking focus of moral blame. In addition, the actions to which the reactive emotions are responses are violations of expectations we accept, and this correctly connects blame with the moral notions of obligation, right, and wrong.

Once blame is understood in terms of the reactive emotions, however, we also have a natural and appealing explanation to hand of what unifies the sanctioning responses to which the stance of holding people responsible disposes us (such as avoidance, censure, denunciation, reproach, and scolding). These can all plausibly be understood as forms of behavior that serve to express the reactive emotions to which we are subject when we blame people for their moral failings ....

For all of these reasons, the reactive account seems a promising interpretation of the stance of holding people morally responsible. But does ... the account yield a way of understanding the practice of holding people responsible as something other than an expression of cruelty and vengeance? It is difficult to answer this question without knowing what cruelty and vengeance consist in, and I do not have a general account of these things to offer; but to start with, it should be noted that not all actions that aim to inflict harm are necessarily cruel—if, for example, I deliberately harm someone who has attacked me, as a strategy of self-defense, what I have done is not cruel. What matters is the goal that the infliction of harm is designed to achieve. Thus we might take as a paradigm of cruelty behavior that aims to inflict suffering (physical or psychic) as a way of subordinating the sufferer to one's will.24 The

resentment others' failure to comply with moral obligations in their relations with oneself. I shall henceforth take these qualifications as understood when I refer either to the susceptibility to reactive emotions or to the belief that those emotions would be appropriate.

24 Compare Judith N. Shklar's definition of cruelty, in Ordinary Vices (Cambridge, MA: Harvard University Press, 1984), p. 8: "the willful inflicting of physical pain on a weaker being in order to cause anguish and fear" (my italics). The mistaken restriction of cruelty to cases involving physical pain is corrected by Shklar on p. 37 of Ordinary Vices, where she acknowledges a species of moral cruelty that involves deliberate humiliation without physical pain. The references to the weakness and humiliation of the victim of cruelty capture the element of subordination and domination made central in my paradigm of cruelty.
worry will then be that the blame and sanctioning responses characteristic of holding people responsible are responses that of their nature approximate to this paradigm, so that they are essentially cruel and vengeful.

No doubt many things have been done in the name of blame and moral sanction that are cruel and vengeful in this way—the history of punishment in the law provides a host of familiar and sobering examples. 25 But the issue is whether approximation to the paradigm of cruelty is essential to the practice of holding people responsible, and the reactive account suggests that it is not. Granted, blame and moral sanction often cause suffering for the person at whom they are directed. But it need not be the case that they aim to cause suffering, as part of a strategy of subordinating the sufferer to one’s will. What is essential to the harmful moral sanctions, on the reactive account, is their function of expressing the emotions of resentment, indignation, and guilt; this is the real point of such responses as avoidance, denunciation, reproach, censure, and the like, and what holds them together as a class. Sanctioning behavior belongs to the syndrome of responses to which the reactive emotions dispose those who are subject to them, because the connection with reactive emotions is part of the conventional meaning of such behavior. 26 We learn the concepts of indignation, resentment, and guilt in part by learning to see their connection to sanctioning behavior, and the adequate expression of those emotions often requires such behavior. Insofar as it plays this expressive role, however, sanctioning behavior would not seem to be essentially cruel or vengeful, for the expressive role does not require the deliberate infliction of suffering as a means to the domination of another; it is one thing to inflict harm with the aim of subordinating a person to one’s will, and quite another to inflict harm with the aim of expressing a moral emotion to which one is subject.

Of course, if the emotions expressed by sanctioning behavior were themselves emotions of blind hatred or anger, involving a desire to inflict harm on their object as an end in itself, the expressive function of that behavior might not rescue it from the charge of cruelty. But the reactive emotions expressed by moral sanctions are not of this kind. Rather, they are focused emotional responses to the violation of moral obligations that we accept. In expressing these emotions, then, we are not just venting feelings of anger and hatred, in the service of an antecedent desire to inflict harm for its own sake; we are demonstrating our commitment to certain moral standards, as regulative of social life. Once this point is grasped, blame and moral sanction can be seen to have a positive, perhaps irreplaceable contribution to make to the constitution and maintenance of moral communities: by giving voice to the reactive emotions, these responses help to articulate, and thereby to affirm and deepen, our commitment to a set of common moral obligations. 27

25 Consider also Nietzsche’s bad conscience and Freud’s unconscious sense of guilt .... These are cases where guilt has a sadomasochistic aspect, being experienced as part of a strategy for inflicting torment on oneself.

26 Perhaps this is what P.F. Strawson means when he writes that a preparedness to acquiesce in the infliction of suffering is “all of a piece” with the reactive emotions (FR, 33–477).

The reactive account thus reveals that holding people responsible need not be wedded to attitudes of cruelty and vengeance. But it continues to tie this stance to emotions that are negative and disapproving, and this raises different questions. For instance, it seems that we hold people morally responsible not only for actions that violate moral obligations we accept, but also for the morally worthy actions they perform. There is a difference between doing something that inadvertently helps another person and deliberately aiming to help another person, and part of the difference seems to be that in the latter case, but not the former, we credit the agent for the good deed in a way that implies responsibility for it. Can this positive conception of moral responsibility be captured within the framework of the “reactive account”?

To answer this question, it will be useful to articulate a hitherto implicit equivocation in the notion of holding a person to a moral obligation (which one accepts). That stance, as presented so far, would seem to admit of a dispositional and an occurrent interpretation. On the occurrent interpretation, to hold a person to a moral obligation is either to be subject to an episode of reactive emotion, because the person has breached some moral obligation that we accept, or to believe that the violation would make it appropriate for one to be subject to such a reactive emotion. Obviously we cannot hope to understand responsibility for morally worthy acts in terms of this occurrent notion alone, because the person held responsible for a worthy action has not violated any moral obligation we accept, and is not believed to have done so. The occurrent reading yields an interpretation of holding a person responsible for some particular moral wrong; it tells us what it is to regard a person as having done something blameworthy.28 But to regard someone as blameworthy for an action is miles away from the idea of responsibility for a morally worthy or admirable performance.

Consider, then, the dispositional reading of holding someone to an obligation. According to this construal, to hold a person to a moral obligation is to be susceptible to reactive emotions in the case that the person breaches the obligation, or to believe the person to be the sort of person whose violation of moral obligations would make it appropriate to be subject to a reactive emotion. This reading of the notion yields an interpretation not of regarding someone as blameworthy for a particular act, but of regarding someone as a morally accountable agent. To hold a person to moral obligations one accepts, in this dispositional sense, is to view the person as the sort of agent whose violation of moral obligations one accepts would render reactive

the law. Of course, legal punishment incurs a higher justificatory burden than the stance of holding people morally responsible, since the harms to which it exposes people are far graver and more systematic—there is a great difference between imprisonment and reproach. Expressive considerations help us to understand the nature of punishment, and perhaps to rebut the suggestion that the institution of punishment is essentially cruel, but they do not by themselves suffice to justify hard treatment of criminal offenders.

28 It is important to bear in mind that what is at issue is moral responsibility for a wrong. Of course I might hold someone legally responsible for an act without either being subject to resentment and indignation or thinking such emotions would be appropriate (by bringing charges against the person in court, for instance). This shows that we can hold people responsible without holding them morally responsible.
emotions appropriate. But clearly one can view people as morally accountable in this way—adopting toward them the dispositional stance of holding them to obligations one accepts—even on occasions when they have done things that satisfy or exceed our moral obligations. This suggests the following account of responsibility for morally worthy actions; to hold a person morally responsible for such an action is (1) to hold the person to moral obligations one accepts, in the dispositional sense, and so to view the person as a morally accountable agent; and (2) to believe the person has done something that meets or exceeds the moral obligations one accepts.

Proceeding in this way, one in effect treats asymmetrically the cases of responsibility for morally worthy and unworthy actions. When we hold a person responsible for an unworthy act, we are subject to a negative reactive emotion because we believe the person to have violated a moral obligation we accept, or we believe that such an emotion would be rendered appropriate by the violation. By contrast, in the case of responsibility for worthy acts, we do not suppose there to be any particular positive sentiment that we are or ought to be subject to. We suppose only that the agent held responsible has done something that meets or exceeds the moral obligations we accept, and that at the time of action she was the sort of person we hold to such obligations, in the way that is dispositionally connected with the negative reactive emotions. This asymmetry in the accounts of the negative and positive cases seems to me to mirror our practice of holding people morally responsible, with its special connection to the negative responses of blame and moral sanction. Holding a person responsible for an unworthy action, or regarding the person as blameworthy because of the action, goes beyond believing the person to have done something morally unworthy in that it is linked with a range of disapproving emotions that hang together as a class (as I argued in the previous chapter). To hold a person responsible for a worthy action, on the other hand, does not seem presumptively connected to any positive emotions in particular. Of course when people exceed our moral demands in ways that benefit us (for instance, by suffering great inconvenience to do us a good turn), we are often subject to feelings of gratitude. But gratitude is not called for in all cases where actions exceed the moral obligations we accept: consider the category of supererogatory acts that do not benefit us in any way. More generally, we hold people responsible for morally worthy acts that do not exceed the moral obligations we accept, but that merely comply with those obligations—acts such as keeping promises, telling the truth, not

---

29 Note that holding someone accountable, on this interpretation, is always relative to some set of moral obligations. A given agent is not held accountable simpliciter, but with respect to a specified set of moral expectations one accepts. It is accordingly possible to hold a person only selectively accountable—accountable with respect to some obligations one accepts, but not others ....

30 It has been pointed out to me that gratitude is sometimes also appropriate in cases where no action that is morally exceptional has been performed—one might feel grateful toward a secretary who has served one dutifully for many years. What one feels grateful for, in a case of this sort, is loyalty or dependability or service over time (which might itself be exceptional).
harming others, and so forth. In these cases it is especially clear that responsibility for worthy acts need not be connected with any distinctive sentiments.31

Even if this point about the morally worthy cases is granted, however, it might still be thought that the reactive account associates responsibility too closely with negative and potentially punitive emotions. Isn’t it possible to hold people morally responsible without being subject to any malicious sentiments toward them when they violate the moral obligations we accept? Gandhi and King have been suggested as interesting cases in this connection, as persons who demanded conformity with important moral principles but foreswore malicious or punitive responses toward those who had flouted such principles in the past.32 Their example seems to tell against the claim that the stance of holding people responsible should be understood in terms of the susceptibility to such emotions as guilt, resentment, and indignation. As I defined them, these emotions are negative attitudes, in that they are forms of disapproval and include a disposition to sanctioning behavior that serves (I have suggested) to express the emotions when they are felt. But Gandhi and King apparently avoided disapproval and the sanctioning behavior that expresses it, while continuing to hold their opponents accountable for the moral wrongs they committed. Can this be understood as a refinement of the stance of holding people morally responsible, consistent with the reactive account of it that I have proposed?

I think it can, if we see the attitude of Gandhi and King toward moral transgressors as one of forgiveness and love. To forgive someone, in the spirit of love, is a complicated stance. It presupposes that one views the person to be forgiven as having done something that would make resentment or indignation a fitting response—one cannot rightly forgive a person for having done something that would not have rendered one of these reactive emotions appropriate in the first place (as with the actions of an infant). Rather, in forgiving people we express our acknowledgment that they have done something that would warrant resentment and blame, but we renounce the responses that we thus acknowledge to be appropriate.33

31 I have tried to account for moral responsibility for dutiful actions. But there is another kind of “deep” responsibility that one may have for such actions, [which we might refer to as] as autonomy. Thus an agent might be responsible for dutiful actions not just in the sense of being a morally accountable agent who complies with moral obligations, but also in the sense that those actions disclose her own values and commitments (her acceptance of the obligations with which she compiles, for instance). It may be that this notion of autonomy yields a more interesting conception of responsibility for positive performances than the notion of moral responsibility.


33 Renunciation is a complicated intentional stance: in renouncing the reactive emotions, we deliberately undertake to rid ourselves of those emotions, or refuse to behave in the ways that ordinarily express them, while recognizing that the conditions that would make the emotions appropriate are nevertheless present. For more detailed discussions of this double aspect of forgiveness, see Aurel Kolnai, “Forgiveness”, as reprinted in Ethics, Value, and Reality: Selected Papers of Aurel Kolnai (Indianapolis: Hackett Publishing Co., 1978),
If this is correct, however, then forgiveness turns out to be a way of holding people morally responsible, according to the reactive account of responsibility. For on that account, it is sufficient for holding a person morally responsible that one believes that the person’s violation of moral obligations would make it appropriate for one to be subject to the reactive emotions.

Ordinarily, forgiveness is a reaction to an acknowledgment of fault on the part of the person who is to be forgiven.³⁴ Where Gandhi and King seem to go well beyond ordinary responses, in the spirit of love, is in their adopting the stance of forgiveness presumptively toward people who have violated moral obligations, independently of whether those at fault have acknowledged wrongdoing. For my purposes here, however, the important point is that this attitude is compatible with their continuing to hold people morally responsible insofar as it includes the belief that violations of moral obligations they accept would render the reactive emotions appropriate. Of course, it would be possible for a moral reformer to try to abandon even this belief—for all I have said, this may well have been the aspiration of the historical Gandhi and King, at least some of the time. On the reactive account, one would have to say about such reformers that they are no longer in the game of holding people morally responsible at all. But this does not seem an implausible line to take. On the contrary, without the belief that violations of moral obligations one accepts would at least render the reactive emotions appropriate, the reformer’s stance toward other moral agents would no longer have any connection with the kinds of deep assessment that distinguish moral responsibility. This point will be explained and defended in the section that follows.

The Reactive Account and Moral Judgment

The reactive approach, as presented so far, is a recognizable development of the account of responsibility found in Strawson’s “Freedom and Resentment”. Strawson there writes: “Only by attending to this range of attitudes [the reactive attitudes] can we recover from the facts as we know them a sense of what we mean, i.e. of all we mean, when, speaking the language of morals, we speak of desert, responsibility, guilt, condemnation, and justice.”³⁵ Picking up on this idea, Gary Watson has taken Strawson to be making the “radical claim” that the reactive attitudes “are constitutive of moral responsibility; to regard oneself or another as responsible just is the proneness to react to them in these kinds of ways under certain conditions.”³⁶

Construed along these lines, Strawson’s own approach appears to have a markedly noncognitivist character. Holding people morally responsible is understood not in

³⁴ See FR, 22–3/63, where this is built into the notion of forgiveness. Compare Jeffrie G. Murphy, “Forgiveness and Resentment”, in Murphy and Hampton, Forgiveness and Mercy, pp. 14–34, especially pp. 24–9.
³⁵ FR, 34–5/78.
of holding people responsible for one to fault on the part of go well beyond the concept of forgiveness, independently of purposes here, their continuing that violations of is appropriate. Of even this belief—historical Gandhi would have to say that people morally are. On the contrary, as would at least toward other moral judgment assessment that I defended in the
devolution of the statement. Strawson's emotive attitudes] can mean, i.e. of all we act, responsibility, Watson has taken as are constitutive possible just is the conditions.36

to have a markedly understood not in

---


38 Thus the account I have offered does not attempt to explain the "special force" of all moral judgments, but only, at most, of those judgments by means of which we blame people. Compare Scanlon, "The Significance of Choice", p. 169, where it is apparently assumed that an expressivist account of the special force of praise and blame would be part of an expressivist account of the special force of moral judgment generally.
that is accounted for by the reactive emotions. Those emotions are essentially backward-looking, being responses to particular violations of moral obligation, and in this respect they capture exactly the attitude characteristic of blame. Thus, I think it would indeed be strange to suppose that one might blame another person without feeling an attitude of indignation or resentment toward the person, or that one might blame oneself without feeling guilt; attempts to communicate blame generally do function, at least in part, to give expression to such attitudes.

Even if it is granted, however, that judgments of blame are essentially (if not exclusively), expressive of reactive emotions, one might still be in doubt as to whether responsibility can be understood more generally in terms of the susceptibility to such emotions. It is possible, after all, to think that someone has failed to meet the moral obligations we endorse, and to hold the person responsible for the failure—in the terms proposed above, to regard the person as having done something blameworthy—without feeling any particular emotion toward the person. Doesn’t this confirm Scanlon’s more basic point that the susceptibility to emotions is not the essential feature in terms of which we are to understand responsibility? In considering this objection, it is important to recall what is involved in holding someone to an expectation. When I introduced this notion above, I explained it in terms of a disjunction: for an agent to hold someone to an expectation is for the agent to be susceptible to the reactive emotions in cases where that expectation is breached, or for the agent to believe that such emotions would be appropriate for him to feel in those cases, and for him to believe that they would then be appropriate because expectations have been breached. It is this second disjunct that now needs to be emphasized. According to the reactive approach, to hold someone morally responsible is to hold the person to moral obligations that one accepts. This notion of holding a person to a moral obligation is in turn understood by reference to a certain range of emotions; but it is not required that we actually feel the relevant emotion in all the cases in which it would be appropriate to do so. All that is required is that we believe that it would be appropriate for us to feel the emotion in those cases, and that what would make it appropriate is the fact that some moral obligation has been breached.  

For example, you may believe that an especially charming colleague who has cheated and lied to you has done something morally wrong, insofar as he has violated a moral obligation not to cheat or lie for personal advantage, and yet you may have trouble working up any resentment or indignation about his case. In a situation of this sort it would perhaps be strange to say that you blame the colleague for what he has done. But you might, all the same, continue to hold him morally responsible—to regard him as having done something blameworthy—and that is allowed by the reactive account as I have presented it.  

---

39 Note that holding a person responsible does not involve the belief that the reactive emotions are required by the person’s violation of moral obligations, or that one ought to be subject to them, but only the weaker belief that it would be appropriate or warranted for one to feel such emotions in response to the violation of those obligations.

40 Compare Scanlon, “The Significance of Choice”, pp. 165–6, on believing a friend’s action blameworthy without feeling moral indignation or disapprobation.
you actually are subject to a reactive emotion, but an emotional response of this sort is not necessarily required for you to hold your colleague morally blameworthy. It suffices for you to believe that indignation or resentment would be fitting responses on your part, and that they would be fitting because the colleague has done something morally wrong. This seems to correspond to our ordinary judgments of moral responsibility quite exactly. Thus if you know that a moral obligation you accept has been breached, and there are no exonerating circumstances that you are aware of, and you still do not believe that the moral response of indignation or resentment would at least be appropriate on your part, then it seems doubtful that you really do hold the colleague morally responsible for his actions or regard him as having done something blameworthy in this case. Of course you might continue to treat the colleague as if he were morally responsible, and engage in sanctioning behavior (for educational or deterrent purposes, say); but...this is not the same as actually holding the colleague morally responsible for what he has done.

In developing the reactive approach in this way, I am in effect exploiting the close connection between reactive emotions and expectations that I was at pains to emphasize in Chapter 2, for this connection is among the things that distinguish the account from cruder noncognitivist theories. Blame is construed essentially in terms of emotions, but the emotions in question are not arbitrary feelings of disapprobation and dislike; rather, they have propositional contents that are fixed by their connection to moral obligations that we accept. Moreover the nature of this connection with moral obligations is such that the reactive emotions are made appropriate by certain kinds of beliefs, about the violation of the moral obligations we hold people to. This connection is what gives the reactive emotions the backward-looking content and focus that is characteristic of the attitude of blame. But it also explains how we can continue to regard a person as having done something blameworthy even when we do not feel the emotions that would be appropriate responses to the person’s behavior: namely, by persisting in the belief that such emotions would be warranted on our part, despite the fact that we happen not to feel them, and that they would be warranted in virtue of the fact that a moral obligation we accept has been violated. Because of this connection between reactive emotions and moral obligations, it seems misleading to refer to the reactive approach to responsibility, as I have developed it, as a distinctively noncognitivist strategy. The approach explicates moral responsibility in terms of our susceptibility to a range of emotions, but these emotions, in their turn, have an essential cognitive aspect that is given by their connection with moral obligations. 41

This raises a question: if it is not really necessary to be subject to reactive emotions for one’s stance to count as holding someone morally responsible, why bring in such

41 My remarks in this paragraph make clear the degree to which my understanding of the reactive attitudes differs from Bennett’s. He writes of “the non-propositional nature of blaming, praising etc. in Strawson’s account: feelings are made central, and are not tied systematically to any propositions about their objects” (“Accountability”, p. 24). On my account, by contrast, the connection between reactive attitudes and obligations can only be understood if we suppose that those attitudes are systematically tied to propositions about their objects.
emotions at all? Could we not understand moral responsibility just as well in terms (say) of the acceptance of moral obligations, and beliefs that such obligations have been violated? On such an account, we might say that to hold someone morally responsible is to be willing to blame the person for violating moral precepts that we ourselves accept for purposes of practical reasoning, deliberation, and public normative discussion, where blame, in turn, simply expresses the belief that those precepts have been breached. But I take it that an account along these lines would not capture what is distinctive about the stance of holding people morally responsible. Blame would be rendered superficial on this account, reduced to a way of describing what an agent has done, and perhaps registering a causal connection between the agent and the action so described. True moral blame, by contrast, is a form of deep assessment, reflecting an attitude toward the agent who has acted wrongly that finds its natural expression in sanctioning behavior (avoidance, denunciation, reproach, censure, and the like). The reactive emotions are needed to explain this attitudinal aspect of true moral blame and to account for its natural connection with sanctioning behavior. And if the reactive emotions are needed to understand the phenomenon of moral blame, they will be equally necessary to make sense of blameworthiness. Thus one can hold a person blameworthy without actually being subject to an episode of reactive emotion; but, as I have argued, blameworthiness does require the belief that some reactive emotion would be appropriate. Without at least this degree of connection with the reactive emotions, we lose the idea that judgments of blameworthiness are forms of deep assessment, and with it the idea that they are ways of holding a person morally responsible.\footnote{42}

It is worth dwelling on this point a while longer, for a proper understanding of it is crucial if one is to appreciate the advantages of the reactive approach. How else might one try to explain moral responsibility for particular actions, and the related judgments of blame and blameworthiness? Scanlon has defended an alternative account that makes no special reference to reactive emotions.\footnote{43} He agrees that judgments of moral responsibility have a force that goes beyond “mere description” of people’s actions, but he contends that this special force cannot be accounted for in terms of what the moral judge is doing in making the judgment (such as expressing a reactive emotion). Instead Scanlon makes the intriguing suggestion that the origin of the special force of judgments of responsibility should be located in the content of those judgments—in “what is claimed about the person judged”.\footnote{44}

Scanlon explains the content of such judgments in terms of a contractualist moral theory. According to this theory, the basic moral obligations are derived from those

\footnote{42} It follows from this that creatures to whom the reactive emotions were completely unfamiliar would not be capable of holding people morally responsible. This seems in line with ordinary thinking about responsibility. Thus, insofar as Mr. Spock (of Star Trek fame) was not susceptible to human emotions, he was depicted as being not quite able to make sense of such human responses as blame, and as not subject to such responses himself. (The question of whether Mr. Spock was completely incapable of such responses is complicated by the fact that he was of partially human ancestry, so that reactive emotions were perhaps not as utterly alien to him as he himself often made out.)

\footnote{43} Scanlon, “The Significance of Choice”, lecture 1, sec. 6 (pp. 167–72).

\footnote{44} Scanlon, “The Significance of Choice”, p. 169.
principles that could not reasonably be rejected by people seeking unforced general agreement on a common set of principles. “What is essential”, he argues: “is that a judgment of moral blame asserts that the way in which an agent decided what to do was not in accord with standards which that agent either accepts or should accept insofar as he or she is concerned to justify his or her actions to others on grounds that they could not reasonably reject. This is description, but given that most people care about the justifiability of their actions to others, it is not mere description.”

Taken literally, this passage suggests that judgments of blame differ from mere descriptions of an action in that they describe the action as lacking a property—namely, accord with principles that could not reasonably be rejected as a basis for contractualist agreement—that people generally want their actions to possess. But as I noted earlier, and as Scanlon himself agrees, this alone does not distinguish moral blame from many other kinds of description (for instance, descriptions of people as handsome or clever), and so it does not seem to capture what is special about moral responsibility. The further element, he suggests, is to be found in the connection of moral blame with reasons and justifications. To blame s, or to judge s morally responsible for what s has done, is at least potentially a way of requesting an explanation or justification from s. Blame is thus set apart from other forms of unwelcome description by its suitability to serve in a system of public codeliberation; it “differs from mere unwelcome description because it calls for a particular kind of response, such as justification, explanation, and admission of fault”. This connection of blame with reasons and justifications points toward the distinctive conditions of responsibility: blame can only be an appropriate response when it is directed at features of a person—such as intentions, actions, or decisions (in contrast to, say, appearance or intelligence)—that are open to assessment in terms of reasons and justifications.

This complicated proposal can be broken down into a plausible claim about the conditions of responsibility and a more dubious suggestion about the special force of judgments of blame. The plausible claim is that blame is an appropriate response only when it is directed at features of a person that are open to assessment in terms of reasons. What is not plausible is the suggestion that the connection with this kind of assessment sets blame apart from other forms of unwelcome description, accounting for its special force. Note, for instance, that beliefs are like intentions in their being explicable by reference to reasons, and hence are appropriate targets of assessment in terms of reasons. Thus we criticize peoples’ political and aesthetic and scientific opinions if they do not seem to us to be well justified. Being concerned with justification in terms of reasons, such criticism could in principle influence the reflections of the person whose opinions are being criticized; it is thus suited for interpersonal exchange in a way that sets it apart from unwelcome description of a person’s appearance or native talents. But criticism of a person’s opinions in

46 What follows is an interpretation of Scanlon’s difficult remarks on pp. 170–72 of “The Significance of Choice”.
47 Scanlon, “The Significance of Choice”, p. 171; for the reference to code liberation, see p. 167.
terms of reasons is normally very different from moral blame for a person’s actions and decisions. Indeed it is different precisely in that it lacks the distinctive force of judgments of moral blame and moral blameworthiness, the connection to attitudes that gives those judgments their special “depth”. 48 It is one thing to criticize a philosopher’s views about causation and quite another thing to blame the philosopher for supporting racist or sexist hiring practices; the difference seems to consist in the fact that moral blame has a quality of opprobrium that is lacking in criticism of beliefs or opinions. Hence we cannot hope to account for this special force solely in terms of the connection of moral blame and responsibility with justification.

There is, in any case, something peculiar in the very idea that we might account for the special force of moral blame in terms of its concern with justification. Doing so leads Scanlon to trace the special force to the conditions that make blame appropriate; in particular, he looks to the fact that moral blame is appropriately directed only at aspects of persons that are susceptible to being influenced by reasons. But if, as Scanlon seems to agree, judgments of moral blame and moral responsibility have a force that goes beyond “mere description”, it is obscure how one could hope to explain this in terms of the conditions that make such judgments appropriate. To try to do so is to look in the wrong place for the force of judgments of blame and responsibility. The right place would seem to be not in the conditions that make moral blame appropriate, but in the condition of the judge who assigns blame and endorses judgments of responsibility.

**Conclusion**

P.F. Strawson’s “Freedom and Resentment” famously situates responsibility in relation to the reactive sentiments and the forms of interpersonal relationship with which those sentiments are connected. In these excerpts from my book *Responsibility and the Moral Sentiments*, I have attempted to develop and to defend one important strand in Strawson’s rich and multifaceted discussion. Specifically, I have argued that the distinctive stance we adopt when we hold people morally accountable for their actions can best be made sense of in terms of the reactive emotions of resentment, guilt, and indignation. The plausibility of this position depends, however, on an improved understanding of the reactive emotions to which the stance of holding people responsible is constitutively connected. I have proposed a more constrained interpretation of these emotions than that suggested in Strawson’s article, emphasizing the structuring role that demands or expectations play in the reactions of resentment, indignation, and guilt.

Strawson himself seems to think that once the reactive emotions are brought into the picture, “pessimistic” or incompatibilist conceptions of responsibility can fairly quickly be seen to be mistaken. In a naturalistic vein, his article suggests that the reactive sentiments are given to us with the fabric of human social life, so that there is no real question of opting out of the practice of moral responsibility in response

---

48 This is not to deny that there are differences in, say, the content and nature of moral and nonmoral justifications. But such differences do not alone account for the special force of opprobrium that seems to attach to moral criticism and blame.
to the (hypothetical) discovery that determinism is true. A different, Wittgensteinian line of argument in “Freedom and Resentment” contends that intelligible questions about our practices can be posed only from within them, so that there is no position from which a global challenge to the whole practice of moral responsibility might be mounted. Finally, Strawson argues that even if we could raise an intelligible global question about our practice of holding people responsible, the only touchstone for answering that question would be the gains and losses that the practice brings for human life; by this pragmatic standard, the connection of responsibility with the valuable forms of relationship to which the reactive sentiments are connected makes it robustly immune to pessimistic challenge.

I am not convinced by these influential parts of Strawson’s larger argument. We should follow Strawson in understanding the stance of holding people responsible in terms of the reactive sentiments. But this interpretation of moral responsibility leaves plenty of room for a coherent and serious pessimistic challenge to the credentials of the practice. The question is not whether the reactive sentiments are inevitable, or whether they contribute to the enrichment of human life. It is whether it would be fair to adopt this stance toward people in a world in which their actions are the deterministic result of prior conditions and events. A Strawsonian interpretation of the practice of holding people responsible sets the terms in which this challenge should be understood. But the ultimate resolution of the issue between optimists and pessimists requires a detailed engagement with normative issues about the fairness of reacting to the things people do with such sentiments as resentment, indignation, and guilt.  

---

49 See the rest of my Responsibility and the Moral Sentiments (Cambridge, MA.: Harvard University Press, 1994), for a defense and development of this interpretation of the traditional debate about moral responsibility and its conditions.