Abstract

Traditionally, the central threat to the defensibility of the range of practices and attitudes constitutive of moral criticism has been seen to be posed by the Causal Thesis, the view that all actions have antecedent causes to which they are linked by causal laws of the kind that govern other events in the universe. In such a world, agents lack the sort of underived origination and agency required for the appropriateness of moral criticism. However, Peter Strawson’s “Freedom and Resentment” marks a move away from a metaphysical conception of agency and conditions of the appropriateness of moral criticism. On Strawson’s account, the problem of moral responsibility is centrally a normative problem, a problem about the moral norms that govern interpersonal relationships, and the conditions of appropriateness of the range of attitudes and sentiments occasioned by the agents’ fulfillment or non-fulfillment of these norms.

In this dissertation I argue that the success of normative conceptions of conditions of appropriateness of moral criticism is contingent of the amelioration of the tension between two strategies in “Freedom and Resentment.” Naturalist interpretations hold that sentiments and practices constitutive of moral criticism are natural features of human psychological constitution, and therefore neither allow nor require justification. Rationalist interpretation, by contrast, are based on an analysis of conditions under which moral criticism can be justifiably modified or suspended.

Both of these strategies, I argue, are false. The naturalistic interpretation is false not because of its inability to offer a plausible account of the conditions of justifiability of reactive attitudes, but rather because of its inability to offer a principled account of the
way moral norms are grounded. The rationalistic interpretation, in turn, not only relies on an implausible psychological account of conditions of responsible agency, but puts an unacceptable emphasis on the agent’s intention. A plausible interpretation of the normative strategy requires emphasizing not only the significance of attitudes and feelings, but also the role reasons play in constituting moral norms and justifying moral criticism.
Acknowledgments

My first philosophical debt in writing this dissertation is to my supervisor and teacher, Professor Rahul Kumar, who patiently read many drafts of this work and others, and offered friendly, helpful, and constructive feedback and advice at every stage. His contributions to this project are too numerous to acknowledge individually. However, a blanket acknowledgment is required: without his guidance, mentorship, and generosity of heart and mind, this project would never have come to completion.

Many others also made significance contributions to this project. The Department of Philosophy at Queen’s University, and its distinguished faculty, made studying philosophy a joy. In particular, I must acknowledge Professor David Bakhurst, who read the entire final draft and made many helpful suggestions. Also, Professor Stephen Leighton, Joshua Mozersky, Deborah Knight, and Alistair MacLeod, with all of whom I have had the privilege of discussing many ideas and issues.

I have presented parts of this work at a number of conferences, including Society of Christian Philosophers, American Philosophical Association, Canadian Philosophical Association, and Western Canadian Philosophical Association. I also presented parts of this work at the Department of Philosophy’s colloquium, as well as the Graduate Colloquium, and I have benefitted greatly from the comments of those who participated.

This project was made possible by generous funding from Queen’s University, as well as a doctoral fellowship from the Social Sciences and Humanities Research Council of Canada.

(July, 2010)
# Table of Contents

Abstract.................................................................................................................................i  
Acknowledgments...........................................................................................................iii  
Chapter 1   Introduction.................................................................................................1  
Chapter 2   The Naturalistic Strategy.........................................................................27  
Chapter 3   Exemptions and the Condition of Moral Sense.................................58  
Chapter 4   Excuses and the Requirement of Intention...........................................89  
Chapter 5   Freedom, Fairness, and Reasons.............................................................118  
Bibliography.....................................................................................................................162
CHAPTER 1: INTRODUCTION

1.1 Moral Criticism and the Threat of the Causal Thesis

Moral criticism is a familiar feature of human interpersonal relationships. We are not indifferent to wrongdoing. A colleague who plagiarizes your latest essay, a driver who refuses to observe red lights and stop signs, and a neighbor whose loud music keeps you up until dawn, all of these agents seem appropriate candidates for moral criticism. We are not only liable to feel attitudes such as indignation and resentment towards these agents, but ostracism and censure also seem appropriate responses to what they have done. Furthermore, self-blame (that is, the self-directed attitudes and practices such as guilt and shame) is a familiar phenomenon of moral life, where an agent is liable to view him or herself as an appropriate candidate for moral criticism in virtue of having committed a wrong.

While the attitudes and practices constitutive of moral criticism are deeply entrenched in everyday life, questions have been raised about their defensibility. Our strong commitment to, and confidence in the justifiability of, the range of attitudes and practices constitutive of moral criticism seems vulnerable to the truth of the Causal Thesis. The Causal thesis claims that “...all our actions have antecedent causes to which they are linked by causal laws of the kind that govern other events in the universe, whether these laws are deterministic or merely probabilistic.” The thought here is that as human beings, we are not merely (if at all) immaterial beings, but exist in a physical

1 Throughout this dissertation, by ‘moral criticism’ I mean moral appraisal.
world, governed by physical laws. In such a world we are subject to the very same laws that govern non-human actions and events.  

It is important to pause and develop a clear account of the threat of the causal thesis. The difficulty here is that a thorough causal explanation of human action seems to preclude any rational, purposive explanation of the same. In such a world:

...just as our explanations of the motions of planets no longer require the existence of prime movers to supplement natural processes, so our actions could, in principle, be explained by a complex neurophysiological theory, without reference to a non-natural self that causes them.

Thus according to what Peter Strawson calls reductive naturalism, “the naturalistic or objective view of human beings and human behavior undermines the validity of moral… reactions and displays moral judgment as no more than a vehicle of illusion.” According to Strawson:

Viewed from one standpoint, the standpoint that we naturally occupy as

---

3 Thus for the purpose of the causal thesis, it does not matter what the content of the laws governing human action is. More specifically, contrary to what the thesis of determinism holds, the threat to appropriateness of moral criticism is not generated by the fact that human action is determined (necessitated, fixed) by the conjunction of prior states of the world and laws of nature. What is crucial is that humans act under the very same constraints that merely natural objects do. For an articulation of the threat of determinism, see Bok, Freedom and Responsibility, p. 3; Laura Waddell Ekstrom, Free Will: A Philosophical Study (Westview Press, 2000) pp. 24-5; Peter van Inwagen, “The Incompatibility of Free Will and Determinism,” Philosophical Studies 27 (1975) pp. 185-99.

4 Dennett, “Mechanism and Responsibility,” p. 159.

5 Bok, op. cit., p. 3. Cf. David Hume, Treatise on Human Nature, David Fate Norton and Mary J. Norton, eds. (Oxford University Press, 2000) at p. 257:

It is universally acknowledged, that the operations of external bodies are necessary, and that in the communication of their motion, in their attraction, and mutual cohesion, there are not the least traces of indifference or liberty. Every object is determined by an absolute fate to a certain degree and direction of its motion, and can no more depart from that precise line, in which it moves, than it can convert itself into an angel, or spirit, or any superior substance. The actions, therefore, of matter are to be regarded as instances of necessary actions; and whatever is in this respect on the same footing with matter, must be acknowledged to be necessary.

He goes on to investigate whether the same is true of “actions of the mind,” i.e. human action.

6 Peter Strawson, Skepticism and Naturalism (Methuen, 1985) p. 40.
social beings, human behavior appears as the object of all those personal and moral reactions, judgments and attitudes to which, as social beings, we are naturally prone; or, to put the same point differently, human actions and human agents appear as the bearers of objective moral properties. But if anyone consistently succeeded in viewing such behavior in “purely objective”…or “purely naturalistic” light, then to him such reactions, judgments and attitudes would be alien.7

The idea which these authors are giving voice to is that if a purely naturalistic explanation pertains to human action, if human action can be exhaustively explained by reference to natural causes, then it seems that the area of human agency is shrunk infinitesimally, that we can no longer ascribe genuine moral agency to agents. Given that a naturalistic explanation of human actions is both necessary and sufficient, the agent himself seems to be completely taken out of the picture, or at least absent from the picture in the important way. While viewed from a subjective point of view, (that is, “from the inside,”) we are convinced that we act freely, once we abstract away from this subjective point of view and view human actions as merely a piece of a larger puzzle, as, in Nagel’s words, “part of the order of nature,” it seems that humans act under the very same constraints that pertain to natural objects (e.g. planetary bodies, billiard balls, etc.).

Thus the truth of the causal thesis threatens the defensibility of practices and attitudes constitutive of moral criticism by producing a sense of loss of independence of the agent from the natural world in which the agent lives.8 More specifically, in such a world human beings lack the up-to-us-ness required for the appropriateness of moral criticism.9 This sense of loss of independence from the natural world in turn diminishes

our confidence in the defensibility of the range of attitudes and practices associated with moral criticism.

1.2 “Freedom and Resentment” and The Quality of Will Thesis

Much of the modern and contemporary literature on conditions of moral responsibility has centered on the requirement of up-to-us-ness, and in particular its bearing on responsible agency. Beginning from the intuitively plausible view that some appropriate analysis of up-to-us-ness is an indispensable component of any plausible articulation of conditions of appropriateness of moral criticism, these accounts proceed to investigate the subsequent question of the kind and degree of up-to-us-ness required for the appropriateness and defensibility of the range of practices and attitudes constitutive of moral criticism.

This approach, however, leaves three related questions unanswered: (1) Is the requirement of up-to-us-ness itself defensible? In other words, can the intuition that moral responsibility requires some kind of control withstand scrutiny? (2) Is the problem of moral responsibility, in the final analysis, a metaphysical problem, that is, a problem which admits of the sort naturalistic treatment suggested by the Causal Thesis? And (3) can we articulate an exclusively normative account of conditions of responsible agency, that is, an account of moral responsibility that is not vulnerable to the metaphysical concerns embodied in the thesis of determinism and the Causal Thesis?

Peter Strawson's landmark essay “Freedom and Resentment”\(^\text{10}\) provides valuable tools and insights for investigating these questions. Strawson characterizes his project as

---


that of reconciling two diametrically opposed views about the compatibility of determinism and practices of holding agents accountable for their actions. The Strawsonian pessimist, who is the traditional incompatibilist, argues that if determinism is true then not only the practices of praise and blame, but also all concepts of moral obligation lose their justification.\footnote{Strawson, “Freedom and Resentment,” p. 59.} By contrast, the Strawsonian optimist espouses a forward-looking, consequentialist conception of the nature and justification of moral criticism, according to which our practices of holding agents accountable for their conduct are justified and rendered appropriate by their ability to regulate conduct in certain socially desirable ways.\footnote{Strawson, op. cit., p. 60. P. Nowell-Smith, “Free Will and Moral Responsibility,” in Mind 57 (1948) pp. 45-61, at p. 56.} Moritz Schlick, for instance, having concluded that the chief task of ethics is the explanation of moral behavior,\footnote{Moritz Schlick, Problems of Ethics (Dover Publishers Inc., 1939), p. 28; cf. J. J. C. Smart, “Free Will, Praise and Blame,” in Gary Watson, ed., Free Will 2nd ed., pp. 58-71.} goes on to argue that our practices of holding agents accountable for their conduct are justified only in so far as they facilitate the regulation of conduct:

Punishment is concerned only with the institution of causes, of motives of conduct, and this alone is its meaning. Punishment is an educative measure, and as such is a means to the formation of motives, which are in part to prevent the wrongdoer from repeating the act...and in part to prevent others from committing a similar act.\footnote{Schlick, op. cit., p. 152. For the same argument see J. J. C. Smart and Bernard Williams, Utilitarianism, for and against (Cambridge University Press, 1973) p. 53.}

Accordingly, on the optimist’s account the focal point of moral criticism is the prevention of further harm, and not, as the retributivist theories claim, any concern about the intrinsic value of punishment, which the offender deserves to receive in virtue of having performed the offensive act.
There are two ways that, on the optimist’s account, this prevention can be achieved. It can be achieved directly by rendering the agent either physically or psychologically incapable of repeating the offending action. Incarceration, for example, serves to immediately render an offender incapable of causing further harm to the general public. But there is also an indirect consequentialist benefit in punishment. Our practices of holding agents responsible for their conduct is a form of applying “moral pressure,” the focus of which is to make a repeat of the offense in question less likely by the exercise of influence over the agent’s motives.\(^\text{15}\) This moral pressure can be exercised either on the agent himself or on third parties. For example, seeing that committing a certain transgression, e.g. a traffic violation, is followed by a range of sanctions (e.g. monetary penalties, loss of driving privileges, social scorn, etc.) makes not only the offender himself less likely to commit the offense again, but also serves to dissuade other agents with similar tendencies from committing similar acts.

Arguably there is something attractive about the optimist’s account. It is in at least partial accord with our ordinary understanding of the function of moral criticism. Considering the case of children’s moral education can be particularly illuminating. Children are not born full members of the moral community.\(^\text{16}\) This is evidenced by the fact that when they do commit acts which are \textit{prima facie} injurious, they are usually exempted from the sort of consequences that would befall an adult for the same transgression. However, part of the process of moral education is equipping children with the cognitive and affective tools necessary for navigating moral life. One way this is


\(^{16}\) Strawson, “Freedom and Resentment,” p. 75.
achieved is by demonstrating to the child the negative consequences his offensive actions
tend to bring about. This can be achieved not only by holding others accountable for their
conduct, but also by impressing upon the child that he himself can be the target of such
practices. Thus our actual moral experiences seem to lend credence to the optimist’s
conception of the regulative role of the practices of holding agents morally accountable.

Additionally, the optimist’s strategy successfully avoids the problems inherent in
the classical compatibilism’s account of moral responsibility. In attempting to resolve the
putative incompatibility between mechanism and the appropriateness of moral criticism,
classical compatibilism has tended to articulate and defend a conditional analysis of
freedom which seemed to be impervious to the truth of mechanism. Alfred Ayer, for
instance, conceding that being responsible for an action requires the ability to do
otherwise, argues that to be able to do otherwise, one must be able to do otherwise if one
had chosen to do so.17 Formally expressed, according to the conditional analysis of free
agency, the proposition:

\[ A \text{ could have done otherwise,} \]

is logically equivalent to

\[ I \text{f A had chosen (or willed, or wanted, etc.) to do otherwise he would have done otherwise.} \]

---

15-23, at p. 22. Bernard Berofsky traces this compatibilist strategy to G. E. Moore. See Bernard Berofsky,
University Press, 2002) pp. 181-201. Robert Kane notes that this analysis was not initially formulated for
the purpose of responding to incompatibilism, but rather as a formulation of the ordinary conception of
freedom. See Robert Kane, “Introduction: The Contours of Contemporary Free Will Debates,” in his
*Oxford Handbook of Free Will* (Oxford University Press, 2002) pp. 3-41, at p. 13. One such ordinary notion
is that of political freedom. John J. Cleary, for example, has argued that the ancient notion of freedom
should be understood as political and social freedom. See John J. Cleary, “Popper on Freedom and Equality
Such a strategy seeks to offer a new analysis of free agency which is compatible with the truth of mechanism. For it seems clear that it is not the case that the truth of mechanism would undermine (2). This is because according to classical compatibilism, freedom in the requisite sense requires only absence of physical constraints.

But as critics have noted, such an analysis fails to fully capture the conditions of free agency. The crucial problem here is that (2) is not logically equivalent to (1). In other words, it is possible that if one had chosen to do otherwise, one would have been able to do otherwise, without it being the case that one could have chosen otherwise, as cases of phobia and compulsive behavior demonstrate.18 In order to complete the analysis, we require a further premise, namely:

\[(3) \quad \text{A could have chosen to do otherwise.}\]

But not only does this seem to make the analysis circular, as we are now in need of a conditional analysis of (3), it also invites incompatibilist criticisms that in a world in which mechanism is true, agents cannot choose to do otherwise. The optimist’s strategy successfully steers clear of this muddle.19

Nevertheless, the optimist’s strategy is flawed, because it fails to capture something essential about the nature of moral responsibility. In particular, the optimist’s framework seems to leave the door open for excessively harsh responses to offensive behavior.20 In particular, if prevention of further harm is, as the optimist claims, the *sole*

---


20 Catherine Rogers, “Retribution, Forgiveness and the Character Creation Theory of Punishment,” in
end of our practices of holding agents to account for their behavior, then it seems that the imposition of excessively harsh penalties on individuals might end up being justified, if it tends to bring about its intended effects. In other words, on the optimist’s account, excessive and disproportional punishment, i.e. “making an example out of someone,” is permissible provided it brings about better consequences, which seems an implausible conclusion.

The optimist can respond that disproportionate punishment is precluded on consequentialist grounds. Excessive punishment is precluded for the simple reason that it would fail to achieve the utilitarian aim of promoting utility. For example, it might be the case that we could greatly reduce the number of sexual offenses committed by imposing an automatic life-time incarceration on all sexual offenders. But, the optimist will argue, the imposition of excessively harsh punishments on sexual offenders will tend to reduce the overall utility, by giving rise to intense sentiments of alienation and resentment on the part of the excessively harshly treated offenders. Thus it seems that the disutility resulting from the excessively harsh punishment handed out outweighs the utility produced, and the disutility avoided, by the elimination of sexual offenses.  

But this response is unsatisfactory. Its central weakness stems from the fact that it merely stipulates that excessive punishment is bound to produce more disutility than
utility. But while this may well be true in some cases, it is not necessarily true in all. The challenge posed against the utilitarian was that if prevention of harm and promotion of utility are the primary ends of punishment, then we lack the resources to rule out severe and intuitively unjust treatment of agents, on the assumption that such severe treatments are successful in achieving the aim of the prevention of further harm. It is no response to this objection to merely deny the assumption of the argument, and stipulate that excessive treatment can never lead to improved utility. In other words, the utilitarian needs an argument to show that excessive punishment would necessarily result in diminished utility, and therefore ruled out on consequentialist grounds.

It is important to note that the problem faced by the consequentialist is not merely that no such argument has been offered to-date. Instead, the consequentialist faces a conceptual challenge. This is because in espousing an entirely consequentialist conception of the nature of moral criticism, the optimist loses sight of the fact that moral criticism is in fact moral assessment of the agent and his conduct. This assessment must therefore necessarily track what the agent has done. Such a resource, however, is unavailable to the optimist, who espouses an entirely forward-looking account of moral criticism.

Considering the failure of the optimist’s account can expose the missing element in both pessimist and optimist’s accounts. Strawson argues that both the optimist and the pessimist commit a common mistake when engaging in philosophy in “our cool, contemporary style,” by losing sight of a central feature of our interpersonal relationships. In a crucial passage in his “Freedom and Resentment,” Strawson

---

22 Strawson, “Freedom and Resentment,” p. 64.
summarizes the basis of ascriptions of responsibility as follows:

We should think of the many different kinds of relationships which we can have with other people — as sharers of common interest; as members of the same family; as colleagues; as friends; as lovers; as chance parties to an enormous range of transactions and encounters. Then we should think in each of these connections in turn, and in other, of the importance we attach to the attitudes and intentions towards us of those who stand in these relationships to us, and of the kinds of reactive attitudes and feelings to which we ourselves are prone. In general, we demand some degree of good will or regard on part of those who stand in these relationships to us, though the forms we require it to take vary widely in different connections. The range and intensity of our reactive attitudes towards goodwill, its absence, or its opposite vary no less widely.23

We can extract three ideas from this passage. First, Strawson calls attention to the wide range of interpersonal relationships we participate in, and the many roles we occupy within such framework. Human beings are essentially “social animals,” so constituted as to have the tendency to associate with other persons and conglomerate into a society.24

Second, supervening on these relationships is a demand for good will, a demand we have towards each other in terms of the quality of each other’s attitudes.25 We demand of each person that he should display appropriate regard for the legitimate claims, entitlements, and interests of others. In order to illustrate this, consider the following example (call it the speeding driver example). Suppose one is about to cross the street with the pedestrian green light when a driver, ignoring the posted speed limit, runs his red light and speeds away. If one had been hit by the car, there would be an easy and straightforward case for the appropriateness of blame. In such a case, blame would be in response to an injury, that is, a concrete harm. However, as luck would have it, one saw

23 Strawson, “Freedom and Resentment,” p. 76.
24 Hume has noted the importance of the combination of these two factors in the tendency of humans to “seek society.” Hume, A Treatise of Human Nature, p. 258.
the car approaching at an excessive speed, and did manage to get out of the way in time. There was, therefore, no injury. But the fact that the driver did not care about the existence of the pedestrian, the fact that he did not care for his safety, the fact that he did not care for how his actions might affect others, in short, the fact that the driver did not show sufficient regard for the pedestrian’s interests and legitimate claims, amounts to a violation of the required good will and reasonable regard. This is because the subject matter of morality is not merely what agents do and do not do, but rather what T. M. Scanlon has called the agent’s judgment sensitive attitudes. These include not only intentions, but also any attitude that is open to rational criticism. Framing the subject matter of the expectation of good will in terms of judgment sensitive attitudes means that merely showing indifference to the interests of another is sufficient for a prima facie violation of the expectation of good will.

I have focused on the question of the nature of the expectation of good will and the ways in which it can be violated. But it is a further question, and the third point implicit in the quotation above, what consequences violations of these expectations entail. According to Strawson, the extent to which agents fulfill or flout the expectations placed on them results in a class of responses which Strawson terms reactive attitudes. These are attitudinal responses persons have to the degree to which agents fulfill the expectations incumbent upon them. If the agent fails to fulfill these expectations, he would be liable to face a set of negative reactive attitudes, such as resentment and indignation. By contrast, if he meets, or more importantly, exceeds the expectations incumbent upon him, he would face a set of position reactive attitudes, such as gratitude.

and approbation.

I want to focus, as I have so far, on the negative reactive attitudes. Consider the speeding driver example. Even if the driver’s actions do not result in injury, the fact that he failed to show appropriate regard for the claims and interests of others evokes certain set of emotional responses. These responses, such as resentment and indignation, are occasioned by the fact that the driver failed to fulfill the demands for due regard incumbent upon him. Like other emotions, reactive attitudes have a distinct phenomenological quality. It feels a certain way to be indignant at someone, to resent him or her, etc..

In this way, reactive attitudes are fundamentally opposed to what Strawson calls the “objective attitude”:

To adopt the objective attitude to another human being [is] to see him, perhaps, as an object of social policy; as a subject for what, in a wide range of sense might be called treatment; as something certainly to be taken account, perhaps precautionary account of; to be managed or handled or cured or trained; perhaps simply to be avoided.27

The objective attitude precludes the wide range of sentiments that are endemic to, and indeed constitutive of, interpersonal relationships.28 Thus to adopt the objective attitude towards someone is to treat him not as an equal participant within the adult interpersonal relationships, but rather as someone who is somehow excluded from it. In other words, we treat him not as a moral agent, but rather as a thing.29

Thus Strawson’s account has two distinguishing features. First, he holds that the

---

problem of moral responsibility is not a metaphysical problem, that is, a problem on which an analysis of up-to-us-ness can bear, but rather a problem about moral norms governing interpersonal relationships. Furthermore, following Hume, Strawson holds that our interpersonal attitudes are doubly-significant. Not only are they significant in that reactive attitudes are familiar features of human interpersonal relationships and indicate something important about the status and quality of our relationships, but also in the antecedent sense that we attach significance to the attitudes agents display towards one another. On Strawson’s account it is the content of a person's character (that is, the range of habits and dispositions peculiar to that person) expressed in his attitudes, that attracts our censure or approval (as embodied in reactive attitudes).  

We can call this the quality of will thesis. According to the quality of will thesis, “morally reactive attitudes are responses to the quality of will expressed in a person’s conduct.” The quality of will thesis provides both necessary and sufficient conditions for the appropriateness of moral criticism. On Strawson’s account moral criticism is appropriate just in case we have reasons to conclude that the agent has failed to show appropriate regard for the well-being of others and the claims they have against us.

But what is the nature of the claims and entitlements that agents have against one another, and under what conditions can we legitimately conclude that an agent has failed to uphold these expectations, such that he or she can be morally criticized in accordance with the quality of will thesis? The question here is not merely epistemic, that is, it is not

---

merely being asked how we can know what the relevant claims are, and when they have been violated. Rather, the questions are centrally normative, namely what are the conditions that are necessary and sufficient for (1) *bona fide* moral constraints on action, and (2) for a warranted conclusion that the agent’s actions, whether or not harmful, are indicative of a morally assessable quality of will?

We can begin to appreciate the importance of these questions by juxtaposing two cases (call this the *detained drunk driver* example). In both cases, I physically restrain someone who has had a bit too much to drink and is therefore not in full control of her motor and psychological faculties, and prevent her from getting into her car and driving away. But in one case, the reason I do this is because I predict, with a good degree of certainty, that if she does drive away, given her lack of general coordination, slow response time, etc., that she will be in an accident. In the other case, the reason I restrain her is not because she is in no condition to drive, but that I simply do not want her to leave, and this merely for some selfish end of mine. Have I, in the first permutation of this case, violated any moral norms? In other words, have I failed to show the requisite degree of good will towards the driver? In order to answer this question, we need an account of the normative foundation of expectations of good will.

There is, in addition to the problem about the content of moral norms, a logically distinct question about conditions under which we can legitimately conclude that the agent has failed to fulfill these norms. The problem here is that it is not the case that every violation of moral norms, as informed by the legitimate claims and relevant entitlements of persons, betrays a morally faulty quality of will that could appropriately
be subject to moral criticism. For example, there are moral norms that forbid the operation of motor vehicles at excessive speeds. Agents who gratuitously violate these norms show contempt for the claims and entitlements of others, and are subject to moral criticism. But suppose an agent is speeding through the city streets not gratuitously (he is not on a joy-ride) but is rather rushing his daughter, who is having a seizure and is blue in the face, to a hospital. In this case, the driver has arguably violated the norms against exceeding the speed limit and has put the well-being of others in jeopardy. But, intuitively, he does not have a faulty quality of will. Resentment and indignation are inappropriate responses to this agent. The question is how do we account for this feature of reactive attitudes.

By the same token, the fact that an agent meets or exceeds his obligations does not necessarily reveal a praiseworthy quality of will. Suppose one late night while driving home, I notice someone whose car has broken down. I stop to help, but do so because I have had a difficult day, and would welcome some unexpected company. In this case, even if I do confer benefits on the stranded person, my actions are not indicative of a praiseworthy quality of will. That is to say, my action, even if helpful to the stranded person, does not reflect well on me, and gratitude would be an inappropriate response to my action. In other words, if the stranded person knows my reasons for helping him, he would reasonably not feel grateful and indebted to me.

The task of providing plausible answers to these two questions is complicated by an hitherto inadequately explored tension between Humean and Kantian tendencies in “Freedom and Resentment.” Commentators have distinguished two competing strategies
in “Freedom and Resentment.” Following Kant, the rationalistic strategy proceeds by articulating a quasi-legal account of counterconditions to moral criticism, and conditions under which reactive attitudes can be justifiably suspended or modified. More specifically, the rationalistic strategy holds that since a thorough renunciation of reactive attitudes would entail a diminished quality of human interpersonal relationships, the decision to suspend reactive attitude can be made only on a limited grounds. This leads to an analysis of the way counterconditions to moral criticism function, with the aim of illustrating that the truth of the Causal Thesis does not warrant excusing all wrongdoing. On Strawson’s account, these counterconditions function by showing either that (1) the agent’s actions are consistent with the fulfillment of expectations made of him and that he has not failed to show appropriate regard for the well-being of others, or (2) that he has failed to fulfill the relevant expectations, but he lacks the relevant moral capacities necessary for full participation in ordinary adult interpersonal relationships, and is therefore an inappropriate candidate for the demands endemic to interpersonal relationships.

Following Hume, Strawson’s naturalistic strategy, by contrast, proceeds by arguing that reactive attitudes are natural features of human psychological constitution, and as a result they neither permit nor require rational justification. The standards of appropriateness of moral criticism are internal to the system of interpersonal transactions. The truth of no theoretical thesis, including the Causal Thesis, can undermine the appropriateness of moral criticism. Thus whereas the rationalistic (i.e. Kantian) strategy

---


17
proceeds by attempting to offer a general justification for reactive attitudes (and, more specifically, conditions of warrant for reactive attitudes) through a systematic analysis of the way counterconditions to moral criticism function, the naturalistic (i.e. Humean) strategy holds that the practices and attitudes constitutive of moral criticism as a whole neither require nor admit of justification.

1.3 Prospectus

The tension between the Humean and Kantian tendencies in “Freedom and Resentment,” and the attendant uncertainty about the precise content of the quality of will thesis, give rise to two related questions. One way to pose this question is to ask whether, in light of the tension identified, we can offer an internally consistent articulation of the quality of will thesis. At the same time, there is a related but more general question that is equally important. This is the question of the implications of the tension between the rationalistic and naturalistic strategies for the project of providing a de-metaphysicalized articulation of preconditions of responsible agency. The question here is, given the tension identified, what shape our de-metaphysicalized account of conditions of appropriateness of moral criticism ought to take?. Does the tension in question entail that the project of providing a non-metaphysical account of conditions of responsible agency is doomed to fail?

These two questions delineate the boundaries of this dissertation. My central thesis is that the examination of the tension between the rationalistic and naturalistic strategies illustrates that the prevailing articulations of the de-metaphysicalized accounts of responsible agency are seriously flawed. Having in place a compelling account of these failures is essential to providing a plausible interpretation of Strawson’s non-
metaphysical account of conditions of appropriateness of moral criticism.

In Chapter Two I will focus on the naturalistic strategy and in particular on the question of whether the natural facts which the naturalistic strategy marshals in neutralizing the threat of the Causal Thesis are capable of providing an adequate basis for the range of practices and attitudes constitutive of moral criticism. I begin, in Section 2.2, by providing a full account of the psychological bases of responsible agency. More specifically, the question here is through what mechanism are the attitudes and practices constitutive of moral criticism generated. In answering this question, I highlight the role the mechanism of sympathy plays in bridging the gap between personal reactive attitudes (that is, resentment towards another on one’s own behalf) and the vicarious, impersonal analogues of resentment (e.g. moral indignation). Just like reactive attitudes, sympathy is a given feature of human psychological constitution, and therefore admits of the same naturalistic treatment.

My second aim in Chapter Two is to argue that the naturalistic strategy is incapable of providing a plausible basis for a de-metaphysicalized account of responsible agency, but that the leading criticisms of the naturalistic strategy have misidentified the source of this failure. I argue that the problem for the naturalistic strategy is not merely that it cannot offer a normatively adequate account of the justificatory grounds of reactive attitudes (as Strawson’s critics have argued), but rather that it lacks the resources to offer a compelling account of the normative basis for the range of demands and expectations on which reactive attitudes supervene. On my account the difficulties for the naturalistic interpretation of the quality of will thesis are generated by the fact that it holds that the
demands for good will and reasonable regard are socially constituted. This social
constitution of the demands for good will, I argue, precludes any deeply normative
justification for demands endemic to interpretation relationships. The naturalistic strategy
therefore leaves mysterious the normative status of the quality of will thesis (and the
norms governing interpersonal relationships).

The failure of the naturalistic strategy motivates consideration of the rationalistic
strategy, and in particular Strawson’s account of conditions under which the attitudes and
practices constitutive of moral criticism can be legitimately modified or suspended. In
Chapter Three I am concerned in particular with the question of whether we can offer a
compelling account of exempting considerations. Exempting considerations do not deny
that the agent has failed to uphold the demands incumbent upon him, but rather seeks to
show that the agent lacks the requisite moral capacities for full participation in inter-
personal relationships, and is therefore an inappropriate candidate for the burden of
demands and expectations associated with participation in normal, adult interpersonal
relationships.

A plausible analysis of this argument requires an account of the nature of
capacities and competences necessary for responsible agency, as well as a deeper
investigation into the very possibility of exemptions. In a recent essay Paul Russell has
defended what he calls the condition of moral sense, according to which the capacity to
feel and understand moral sentiments (such as resentment, gratitude and indignation) is
necessary for being an appropriate target for moral criticism.33 According to Russell,

287-305; cf. John Martin Fischer and Mark Ravizza, Responsibility and Control (Cambridge University
Press, 1998). On Fischer and Ravizza’s account, viewing oneself as an appropriate target for reactive
attitudes is a condition of ownership of the action.
agents who lack the faculty of moral sense therefore lack the capacity to grasp and apply moral norms, and as such are inappropriate candidate for moral criticism. Russell’s defense of the condition of moral sense gives rise to three questions. The first question is whether the capacity to feel and understand reactive attitudes is necessary for the capacity to grasp and apply moral norms. The second question is whether the capacity to feel and understand reactive attitudes is directly necessary for appropriateness of moral criticism. The third question is whether the capacity to grasp and apply moral norms is itself a necessary condition of responsible agency.

The first question is primarily an empirical question, and will not be of particular interest. I will be particularly interested in the second and third questions. In Section 3.4 I focus on the second question, that is, the question of whether the capacity to feel and understanding reactive attitudes is a necessary condition of appropriateness of moral criticism. I argue that even if we grant that moral competence is a necessary component of a plausible account of responsible agency, and that the absence of such a capacity is a sufficient condition for the applicability of exemptions and a warranted suspension of moral criticism, the capacity to feel and understand reactive attitudes is not one such capacity. I illustrate this by arguing that a plausible account of the function of reactive attitudes precludes the claim that the absence of the faculty of moral sense renders moral criticism inappropriate.

As I have just noted, the argument of Section 3.4 assumes that some appropriately analyzed notion of moral capacity is indispensable to a normatively adequate account of conditions of responsible agency. This assumption has received little critical attention.
Traditionally the capacity to understand moral norms has been taken to be a necessary condition of blameworthiness. According to one such account, implicit in moral criticism is a demand that the agent comply with moral norms, a demand which is unintelligible if the agent lacks the capacity to understand what the applicable norms are. The second question in which I will be interested in Chapter Three is whether this assumption is correct, that is, whether some relevant analysis of moral capacity is a necessary component of a plausible account of responsible agency. In Section 3.5 I argue that the content of moral address, and therefore the conditions of its intelligibility, have been largely misunderstood. Arriving at a correct account of the content of moral address suggests that the intelligibility of moral address, and therefore its appropriateness, are not undermined by the agent’s incapacity to understand the nature of moral norms. It follows that the inability of the agent to understand moral norms does not render moral criticism inappropriate.

Two implications of this conclusion are worth noting. First, even if Russell is correct in claiming that agents who lack the faculty of moral sense therefore lack the capacity to grasp and apply moral norms, he turns out to be wrong in the further claim that such agents are inappropriate candidates for moral criticism. Furthermore, and this is the second conclusion, we must shift our focus away from the psychology of responsible agency to the content of moral criticism and its significance of interpersonal relationships. Such a shift, I argue, is necessary for a plausible account of the very possibility of responsible agency.

Having argued in Chapter Three that the prevailing account of exempting
considerations is false, I turn in Chapter Four to examining excusing considerations. Excuses do not deny that the agent possesses the capacities required for full participation in interpersonal relationships, and that he is therefore an appropriate candidate for the range of demands and expectations endemic to these relationships. Instead, excuses deny that the agent satisfies conditions of attributability of some putative harm. Does this argument provide a plausible basis for neutralizing the threat of The Causal Thesis? Answering this question requires an account of conditions of attributability of moral responsibility. According to one prevailing view, advanced by R. Jay Wallace, the question of the moral content of the agent’s will (that is, the question of whether the agent’s attitudes are open to moral assessment), and therefore the attributability of some harm to him are, in the final analysis, questions about the intention with which the agent performs the action.\(^{34}\) My central aim in Chapter Four is to argue that this interpretation of the quality of will thesis is unsuccessful. This argument will proceed in two stages. I will begin by arguing that control (and, more specifically, intention) is essentially a causal concept. I then argue that, as a result, the voluntarist accounts are incapable of capturing the concern of the quality of will thesis with the content of the agents’ attitudes. I illustrate this through an analysis of cases of negligence and omission, and argue that Wallace’s voluntarist interpretation of the quality of will thesis is incapable of providing a compelling account of these. I will attribute this failure to the tension between the quality of will thesis on the one hand and the requirement of control on the other.

In spite of the compelling arguments advanced against the requirement of

intention (choice, control), these conditions continue to enjoy significant intuitive and philosophical appeal. In Chapter Five I pursue diagnostic and remedial aims. My diagnostic aim is to provide an analysis of the plausibility and appeal of the requirement of control. My second aim is to articulate and defend an account of moral criticism that does not appeal to the condition of control. In order to achieve these aims, I will begin by providing, following T. M. Scanlon, a contractualist account of moral criticism. According to Scanlon, the obligations we have to each other arise from the value of persons as rational beings. Moral criticism, on this account, is warranted just in case the agent fails to uphold these reasons. For this reason, moral criticism has serious implications for interpersonal relationships, because it is symptomatic of the impairment of the agent’s relationship with others. On my reading, the crucial claim in Scanlon’s account is that no requirement of up-to-us-ness is necessary for the appropriateness of moral criticism, and in particular our concern with the quality of interpersonal relationships. Furthermore, and relatedly, the truth of the Causal Thesis does not bear on the question of whether there are reasons that flow from the value of persons, nor on the question of whether one’s relationship with others is impaired as a result of one’s failure to adhere to these reasons.

Wallace has purported to show that Scanlon fails to divorce the question of conditions of appropriateness of moral criticism from control. He focuses on the case of wayward desires (desires that resist the agent’s attempts to modify or withdraw them) and argues that moral criticism of the agent for such attitudes is unfair. The claim crucial to Wallace’s argument is that some appropriate analysis of free action is necessary for the

fairness of moral criticism. In Section 5.3 I argue that this claim of unfairness can be understood in either of two ways. On one understanding, to morally criticize someone for a wayward desire is to misidentify the attitude that he does in fact hold. On the other understanding, to morally criticize someone for a wayward desire is to impose an unfair burden on him. In Sections 5.4 and 5.5 I argue that neither of these two construals of Wallace’s objection is successful.

The argument of this dissertation can be construed as what Manuel Vargas has called a Strawsonian-revisionist theory of moral responsibility. Vargas argues that the traditional articulations of Strawsonian compatibilism lack the resources to offer convincing answers to the traditional incompatibilist concerns. These accounts must therefore be modified in order to yield plausible accounts of conditions of responsible agency. It is not my intention to offer an analysis of Vargas’ proposal. It is sufficient for my purposes to note that the arguments of this dissertation illustrate that there are important conceptual difficulties with the traditional interpretations of Strawson. Addressing these difficulties requires a re-interpretation of the normative and naturalistic bases of Strawsonian compatibilism.

This reinterpretation requires the rejection of the naturalistic interpretation of the quality of will thesis, and particular the rejection of thought that conditions of appropriateness of moral criticism can be located within a naturalistic understanding of reactive attitudes and the demands on which they supervene. We must likewise reject Strawson’s view that the possession of moral competence is a necessary condition of

---

responsible agency. In a similar manner, we must reject the voluntarist elements of “Freedom and Resentment,” and in particular Strawson’s insistence that intention is the primary attitude with which moral criticism is concerned, as well as the attendant notions of free agency and control too which Strawson attaches significance.

In spite of these failures, there are significant insights in “Freedom and Resentment.” On my reading, the significance we attach to the attitudes agents display, and the force of moral criticism, rest on substantive considerations about reasons that are at stake. The key to resolving the tension between the Humean and Kantian tendencies in “Freedom and Resentment,” as well as to articulating a thoroughly de-metaphysicalized conception of conditions of responsible agency, lies in discounting the significance of moral psychology and brute facts about human psychological constitution, as well as the concern with agential control. We must instead incorporate into Strawson’s account of reactive attitudes considerations about the role reasons play in governing our attributions of moral responsibility.
CHAPTER 2: THE NATURALISTIC STRATEGY

2.1 Introduction

In this chapter I pursue two aims. My first aim is to articulate a comprehensive account of Strawson’s naturalistic strategy. My second aim is to explore the question of whether the naturalistic strategy is capable of providing a normatively adequate basis for the range of attitudes and practices associated with moral criticism. Critics of Strawson’s naturalistic strategy have tended to locate the failure of the naturalistic strategy in its inability to provide a non-arbitrary account of conditions of justifiability of reactive attitudes. In Section 2.3 I argue that this objection is not decisive against Strawson, for the reason that the fact that the naturalistic strategy takes reactive attitudes to be a natural feature of human psychological constitution is compatible with there existing standards which govern the warranted deployment of reactive attitudes and assignment of blame.

In Section 2.4 I argue that the problem with the naturalistic strategy is that it fails to offer a compelling account of the normative basis of expectations of good will, that is, an account of the connection between the socially-constituted expectations of good will and the bona fide moral norms the violation of which would attract warranted moral criticism. It is in this connection that the failure of the naturalistic strategy becomes clear.

2.2 The Naturalistic Strategy

According to Michael McKenna and Paul Russell, the crucial feature of Strawson’s naturalistic strategy is:

...turning away from conceptual issues about the analysis of “freedom” and “responsibility” and taking a closer look at what actually goes on when we hold a person responsible. That is to say, [Strawson’s] methodology depends less on conceptual analysis and more on a descriptive account of
actual human moral psychology.¹

This descriptive analysis consists of two components. First, Strawson argues that given the sort of beings we are, a sustained and thorough objectivity of attitude is not something we are capable of. According to Strawson:

The human commitment to participation in ordinary inter-personal relationships is, I think, too thoroughgoing and deeply rooted for us to take seriously the thought that a general theoretical conviction might so change our world that, in it, there were no such things as inter-personal relationships as we normally understand them; and being involved in inter-personal relationships as we normally understand them precisely is being exposed to the range of reactive attitudes and feelings that is in question.²

This argument proceeds by observations about the necessarily social nature of human life.³ The claim here is not merely that it is a fact about humans that we tend to form societal associations. There is a deeper claim involved, namely that being in reciprocal relationships with other human beings is a necessary condition of a full and satisfying human life.⁴ The crucial point here is that to be human (in both the descriptive sense and the normative sense just enunciated) is to be involved in interpersonal relationships, to connect with other human beings as objects of a care and concern, as agents capable of wronging and being wronged (or, as the case may be, benefiting and being benefited).

In turn, and as a result, being involved in inter-personal relationships is characterized by a dual vulnerability. First, our associations render us vulnerable to what

⁴ Strawson, op. cit., p. 70; McKenna, op. cit., p. 167.
others think about us, and the attitudes and feelings that accompany this. We are vulnerable to being resented or held in esteem, hated or loved, etc.. In this way, what others think of us is an important component of our interpersonal relationships. Second, we ourselves are liable to feel and experience a range of emotions and sentiments as a result of our involvement in interpersonal relationships. To be involved in a friendship, for example, is to render oneself vulnerable to being betrayed. If a friend fails to honor a promise, or render assistance in an important project, without a compelling excuse, we are liable to experience a range of reactive attitudes and feelings. The same point applies more generally to any sort of social interaction. A stranger on the bus who fails to adhere to the applicable protocols (for instance, by listening to loud and obnoxious music, failing to respect others’ “personal space,” having poor hygiene, etc.) is likely to elicit a set of reactive attitudes and feelings.

The crucial point is that these susceptibilities are endemic to interpersonal relationships, and indispensable features of our humanity. Reactive attitudes play such a fundamental role in daily interpersonal relationships that we simply cannot, as a practical matter, sustain an objective attitude towards others for long. And when we do suspend our reactive attitudes, this is never because of a belief in some theory about the causal history of action. The belief in, or conviction of, the truth of the Causal Thesis is of the wrong logical form to have any bearing on our practices of praise and blame. This is because the questions of the justification of ascriptions of responsibility are internal to the system of practices of responsibility. This system, in Strawson’s words, neither calls for,

---

nor allows, external justification. While certain considerations internal to our practices of holding agents accountable for their conduct might well defeat ascriptions of responsibility under certain considerations, the system of ascriptions of responsibility as a whole does not admit, or even call for, external justification.

It is worth pausing for a moment to make the distinction between internal and external considerations clear. Consider the example of a person who is abused by her spouse. Whatever considerations that might be offered by, or on behalf of, the abusive spouse are internal to the system of reactive attitudes. The plea, for example, that the offending spouse has been under extraordinary pressure at work and therefore has been acting out of character operates from within the system of inter-personal relationships. It begins from an already-established and accepted account of conditions of appropriateness of moral criticism, and suggest that the offender should be forgiven on such grounds. But the question of the consequences of the truth of the Causal Thesis is external to the framework of practices of holding agents responsible. The plea that the abusive husband ought to be forgiven does not begin from an accepted understanding of the nature of ascriptions of responsibility. Our de facto moral practices do not indicate that the truth of the Causal Thesis can serve as a countercondition to moral criticism. We never say that we ought to modify our reactive attitudes towards the husband because the Causal Thesis is true. Such a consideration plays no part in our considered moral convictions.

But what about reactive attitudes endows them with the distinct normative and psychological force necessary for the success of the naturalistic strategy? What allows

---

reactive attitudes to play such a central role in interpersonal relationships? These questions are motivated by a tension between two classes of reactive attitudes. As I have argued above (Section 1.2) Strawson emphasizes the significance of a set of attitudes and emotions that we have to the degree to which others show appropriate regard for our legitimate claims and entitlements. Strawson distinguishes these personal (or participant) reactive attitudes from their impersonal or detached kind. 8 According to Strawson, these vicarious analogues of resentment arise not in reaction to personal affronts, but rather on behalf of another. Consider, again, the speeding driver example. While the victim of the offending driver is likely to feel a set of reactive attitudes, those who witness the event are also likely to be indignant at the driver for his carelessness. It is the latter, i.e. the impersonal reactive attitudes, that Strawson identifies with moral responsibility:

Because of this impersonal or vicarious character, we give them different names. Thus one who experiences the vicarious analogue of resentment is said to be indignant or disapproving, or morally indignant or disapproving. What we have here is, as it were, resentment on behalf of another, when one’s own interest and dignity are not involved; and it is this impersonal or vicarious character of the attitude...which entitle[s] it to the qualification ‘moral.’ 9

According to Strawson, therefore, moral responsibility is constituted by impersonal reactive attitudes. To be morally responsible is for one to be an appropriate candidate for a certain sort of sentiment, namely a socially constituted, vicarious analogue of resentment, “resentment-on-behalf-of-another.” The crucial move for Strawson, therefore, is to highlight the centrality of personal reactive attitudes, and then note our ability to transcend our own view point, and be liable to feel impersonal reactive attitudes, which

8 Strawson, “Freedom and Resentment,” p. 70.
9 Strawson, op. cit., pp. 70-1.
are constituted by an abstraction from their personal counterparts.

There are two questions that we may ask about the identification of moral criticism with impersonal reactive attitudes. The first is the question of the need for such an abstraction away from personal reactive attitudes to impersonal reactive attitudes. For if our aim is to hold agents accountable for their conduct, it seems that personal reactive attitudes, such as resentment, are well capable of doing so. So the victim in the speeding driver example can hold the offender accountable for his conduct by adopting towards him a certain attitude, namely resentment. We therefore need to motivate the abstraction from personal reactive attitudes towards their vicarious counterparts. In other words, we must explain precisely what is missing from the attached variety of reactive attitudes as to render it incapable supporting the weight of ascriptions of responsibility.

As we have seen, the problem of responsibility is a social phenomenon, centrally a problem of how agents, within a moral community, interact with and react to one another. In order to get an accurate picture of the nature of moral criticism, we must examine moral criticism not in isolation, but rather within the context of the system of interpersonal interactions. There is, therefore, an intimate connection between the social nature of practices of holding agents responsible for their actions on the one hand and the impartiality demanded by such practices on the other. According to this view, in deciding questions of moral responsibility, we ought to abstract away from individual idiosyncratic responses to perceived affronts, and instead adopt an impersonal perspective. Thus Hume has noted that:

...every particular man has a peculiar position with regard to others; and it is impossible we could ever converse together on any reasonable terms, were each of us to consider characters and persons, only as they appear from his peculiar point of view. In order, therefore, to prevent those continual contradictions, and arrive at a more stable judgment of things, we fix on some steady and general points of view; and always, in our thoughts, place ourselves in them, whatever may be our present situation.  

According to Hume, therefore, personal reactive attitudes are indicative of each person’s particular position in life, his particular commitments, interests, and projects. But assessments from a particular point of view do not amount to moral assessment of the agents. This requires intersubjective standards of conduct and the adoption of a disinterested view, which is achieved by the abstraction from personal to impersonal reactive attitudes and the identification of moral responsibility with the latter.

Reconsider, for example, the case of the detained drunk driver, where I restrain someone who has had too much to drink and is therefore in no condition to drive, and do so out of concern for her well-being. We can imagine that the driver would be extremely upset and resentful at my preventing her from driving. She might object that she is in a perfectly fine condition, that she has driven under the same conditions before without any negative outcome, and so forth. More importantly, she may object that I am not displaying the requisite respect for and deference to her autonomy. The same attitudes might continue even after she regains her faculties. If moral criticism were constituted by personal reactive attitudes, if conditions of appropriateness of reactive attitudes were constituted by personal standards of resentment, then her reaction would have meant that I was an appropriate target for moral criticism, when this is arguably not the case.

The adoption of an impartial evaluative standpoint, therefore, prevents moral criticism from becoming an idiosyncratic expression of personal taste and preference. Attached reactive attitudes are an expression of what Hume has called self-love, that is, an agent’s inclination to care for and advance his own interests. But self-love, according to Hume, cannot serve as a foundation for morals because “the notion of morals implies some sentiment common to all mankind, which recommends the same object to general approbation, and makes every man, or most men, agree in the same opinion or decision concerning it.” It is our concern for general interest of the humankind, and not merely selfish interests, that motivates our concern with morality. Capturing this component of morality requires the stability of an impartial point of view, which the abstraction provides.

The second question we may ask about the abstraction from personal to impersonal reactive attitudes concerns the manner in which this abstraction is carried out. This inquiry into the manner of abstraction is doubly-necessitated. In the first place, concerns might be expressed about a putative tension within Strawson’s account of the nature of moral criticism. As I have argued, on Strawson’s account moral criticism demands an abstraction away from personal reactive attitudes and the adoption of an impartial standpoint. At the same time, our practices of responsibility require a process of double-identification. We must not only identify with the putative victim and share his or her pain or pleasure, but we must also share the putative offender’s perspective. But such a process of identification, it might be argued, precludes impartiality, or at least pulls

14 Thomas Nagel, *The View from Nowhere* (Oxford University Press, 1986) p. 120.
in the opposite direction. The very requirement of abstracting away from personal perspective seems to undermine our ability to identify with the individual concerns and point of view. The challenge here is to bridge this gap, that is, allow for the identification with the agent’s point of view, without thereby making moral criticism contingent on the agent’s idiosyncratic projects and concerns.

At the same time, and this is the second point, the process of identification ought to preserve and account for the distinctly social aspect of ascriptions of responsibility. In so far as practices constitutive of moral criticism cannot be examined or criticized individually, but must rather be studied within the broader context of social practices, the account of abstraction must show precisely what allows for the distinctly social, communal nature of such practices. In particular, Strawson emphasizes that reactive attitudes, in their personal, impersonal, and self-directed varieties, have “common roots in our human nature and our membership of human communities.”\textsuperscript{15} The question is precisely through what mechanism such a communal nature of moral criticism is sustained.

According to Strawson, the answer to this question can be located in a class of distinctly human feelings.\textsuperscript{16} In particular, he highlights “the central importance of that sense of sympathy, and of a \textit{common} humanity, which underlies not only my indignation on another’s behalf but also my indignation on my own.”\textsuperscript{17} According to Strawson, therefore, what binds the personal and impersonal reactive attitudes, that is, what facilitates the transcendence of the merely personal concerns, facilitating the move away

\textsuperscript{15} Strawson, “Freedom and Resentment,” p. 72.
\textsuperscript{16} Strawson, op. cit., p. 71.
from it and towards the impersonal point of view, is sympathy.

Hume is particularly known for emphasizing the moral dimension of certain human feelings, placing a crucial importance on sympathy. He characterizes sympathy as follows:

No quality of human nature is more remarkable, both in itself and in its consequences, than that propensity we have to sympathize with others, and to receive by communication their inclinations and sentiments, however different from, or even contrary to our own.  

It is beyond the scope of the present discussion to offer a detailed account of the notion of sympathy and its role in Hume’s project. For our purposes it will be sufficient to note that for Hume, sympathy is a mechanism through which one becomes convinced that other persons are similar to one in their actions and feelings, and subsequently shares in their experiences. In other words, our interactions with other human beings lead us to the conclusion that we are, for the most part, similar in physical constitution, as well as psychological dispositions. This has the consequence of enabling us to sympathize with other humans, to share in their pain and joy. It is this human capability to sympathize

with one another, to transcend the self and connect with the concerns of others, that bridges the gap between personal and impersonal reactive attitudes, by enabling us to adopt an impartial evaluative view, and, at the same time, to share in the experiences of others.\textsuperscript{21} It is sympathy that enables the faculty of moral sense, that is, the faculty of distinguishing between morally good and morally bad actions and responding to them accordingly.

The crucial point for Hume, as well as for Strawson, is that there is a natural, indeed irrevocable communal character to human living. Thus Hume notes that it is a natural fact that humans tend to seek society, and this on explicable biological and psychological principles. Endemic to such a communal experience, as we have seen, are expectations we have from our fellow human beings to take into account our interests and legitimate claims. And while we are disposed to respond to personal affronts, we also have a natural ability to commiserate with our fellow human beings. Such commiseration finds a natural expression in our moral practices of holding agents accountable for their conduct through the expressions of reactive attitudes.\textsuperscript{22} In this way, morality is, in Hume’s words, felt, rather than judged.\textsuperscript{23} These reactive attitudes are, as a result, \textit{given} features of our moral experience, that is, they require no general justification.\textsuperscript{24}

Thus what Strawson calls non-reductive naturalism,\textsuperscript{25} does \textit{not} seek to counter the threat of mechanism by argumentation. According to this view,

\textsuperscript{22}Strawson, \textit{Skepticism and Naturalism}, p. 39.
\textsuperscript{23}Hume, \textit{Treatise Concerning Human Nature}, p. 302
\textsuperscript{25}Strawson, \textit{Skepticism and Naturalism}, p. 40.
Such attempts at counter-argument are misguided; and not merely because they are unsuccessful or unintelligible. They are misguided also for the reasons for which counter-arguments to other forms of skepticism have been seen to be misguided; i.e. because the arguments they are directed against are totally inefficacious. We can no more be reasoned out of our proneness to personal and moral reactive attitudes in general than we can be reasoned out of our belief in the existence of body.  

According to non-reductive naturalism, we are so constituted as to hold agents accountable for their actions. This practice is constituted by reactive attitudes and founded on the socially formed expectation of good will, and enabled by the naturally occurring mechanism of sympathy. To attempt to reason the skeptic or the pessimist out of his skepticism or pessimism is not only to misunderstand the nature of the problem of accountability, but also to grant the pessimist or the skeptic too much. The proper response would not be to demonstrate the compatibility of our ascriptions of responsibility with the truth of determinism, but rather to question the relevance of the latter to the former altogether.

Having in place this account of the reactive attitudes and their basis in human psychology, the question I turn to is whether this strategy is capable of neutralizing the threat of mechanism.

2.3 The Naturalistic Strategy and the Justifiability of Reactive Attitudes

As I have already noted, on Strawson’s account moral criticism is constituted by attitudinal reactions agents within a moral community have to the fulfillment or non-fulfillment of certain expectations endemic to interpersonal transactions. Noting this feature of Strawson’s account, Kevin Magill writes:

The shift in focus from beliefs to attitudes in Strawson’s discussion of moral responsibility represents an important challenge to the way the traditional debate has been conducted. If blaming people and holding them responsible were straightforwardly based on beliefs or judgments, it would always be possible to ask what ontological or metaphysical conditions are required for such beliefs or judgments to be true or warranted and to press the inquiry to the point where determinism (or, for that matter, personal identity, or agency, or whatever) becomes an issue. Attitudes, by contrast, are neither true nor false, and are not warranted by anything over and above their standard conditions of applicability. The shift from beliefs to attitudes as the focus of discussion therefore undercuts the characteristic assumption of the traditional antagonists that the problem is about what justifies us in treating people as responsible for their actions.27

According to Magill, if moral responsibility is taken to be constituted by emotional reactions we have to other agents, then it might seem that a crucial cognitive element is absent. In other words, on this view practices of responsibility are merely expressions of attitudes and sentiments of either the offended individual (in the case of personal reactive attitudes) or the society as a whole (in the case of impersonal reactive attitudes), and do not illuminate, or have any interesting relation to, the question of whether the agent is in fact one who is blameworthy. This has led Jonathan Bennett to argue that on Strawson’s account:

Feelings are made central, and not tied systematically to any propositions about their objects. My feeling of indignation at what you have done is not a perception of your objective blameworthiness, nor is it demanded of me by such a perception. It expresses my emotional make-up, rather than reflecting my ability to recognize a blame-meriting person when I see one.28

Thus according to what Watson calls Strawson’s expressivist theory of moral responsibility,29 reactive attitudes constitutive of moral criticism might seem to be non-

29 Gary Watson, “Responsibility and the Limits of Evil, Variations on a Strawsonian Theme,” in Ferdinand
propositional in nature. This, in turn, threatens to sever any necessary connection between moral criticism and the objects thereof. In other words, according to this charge there is an absence of a necessary connection between on the one hand practices of holding agents responsible, constituted by reactive attitudes of both personal and impersonal variety, and on the other an assessment of what the agent has done. If, as Strawson claims, reactive attitudes exhaust the concepts and practices of moral responsibility, one might legitimately wonder the extent to which considerations about the agent and his actual moral appraisability necessarily figure into Strawson’s theory.

The most natural way to understand the objections posed by Magill and Bennett is as a charge of non-cognitivism against the naturalist interpretation of the quality of will thesis. Understood this way, the naturalistic strategy is accused of severing any necessary connection between moral criticism on the one hand and the objects thereof on the other. According to this objection, the appeal to reactive attitudes as given facts about human psychological constitution precludes any questions about the standards of the warrant of moral criticism. This charge would be particularly embarrassing for the naturalistic strategy because it suggests that it is no better than the consequentialist accounts of justification of punishment. Recall that the central objection against the optimist’s account of moral criticism was that it failed to provide a plausible account of what the agent has done, and instead focused on the ability of punishment (and, more generally,

31 Strawson argues “Only by attending to this range of 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.” “Freedom and Resentment,” p. 78, my emphasis.
more criticism) to regulate conduct in socially desirable ways. The charge of expressivism against the naturalistic strategy likewise holds that if we accept the naturalist interpretation’s account of reactive attitudes as given facts about human psychological constitution, we would lack the resources to provide plausible answers to the question of the justifiability of moral criticism.

However, this charge is not successful against the naturalist interpretation. This is evident in the normative interpretation of the quality of will thesis. The relevant question for Strawson is centrally normative, namely under what conditions would it be appropriate to hold an agent morally responsible, that is, what conditions would (and ought to) inhibit the stance of holding agents morally responsible?. Therefore, Strawson is not to be interpreted as making the appropriateness of moral criticism contingent solely and exclusively on the attitudes of agents within the moral community. Rather, according to McKenna, moral emotions themselves are reactions to the fulfillment (or lack thereof) of the sense of entitlement agents have towards one another, in particular in terms of whether others’ actions uphold the expectation and requirement of respect each agent has against the other. It is in this dynamic interaction between, on the one hand, the expectation and demand we have of others vis-à-vis the quality of their will towards us, a sense of entitlement we feel, and on the other the concrete reactions we have to fulfillment or the lack thereof of these expectations, i.e. the moral emotions, that moral responsibility consists. And it is in this sense that Strawson’s conception of moral

---

responsibility is not to be cashed out in terms solely of the stance others take towards the agent, but that these two aspects of moral responsibility (i.e. the expectation of good will and moral reactive attitudes) are “mutually supportive.” 34

Therefore, on an initially plausible reading, Strawson, far from identifying being an appropriate candidate for moral criticism with merely being subject to certain range of attitudes, places the following constraint on ascriptions of responsibility: “Being morally responsible and legitimately holding morally responsible are to be settled exclusively in terms of the moral quality of the will with which an agent acts.” 35 This constraint, according to McKenna, provides a normative standard against which ascriptions of responsibility can be measured. This neutralizes the incompatibilist objection that Strawson’s account of reactive attitudes is unable to capture the essentially normative nature of ascriptions of responsibility, and in particular the concerns the libertarian is bound to raise with respect to the alleged inability of Strawson’s account of reactive attitudes to capture the relevant capacities essential to moral responsibility.

What is important to emphasize is that from the fact that reactive attitudes and expectations of good will are, on the naturalist interpretation, given facts about human psychological constitution, it does not follow that there are no standards which can be invoked to assess the warrant of reactive attitudes. For example, it is a live and interesting question whether a father who is rushing a family member to the hospital and in the course of doing so drives at excessive speeds is indeed an appropriate target for resentment and moral indignation. This question is not rendered moot if we accept the

naturalist interpretation’s account of reactive attitudes as given features of human psychology. In other words, from the fact that, on the naturalist interpretation standards of moral criticism are internal to the system of interpersonal moral transactions, it does not follow that there are no standards at all.

It may be objected that this argument is deficient in one crucial respect. The thought here is that my construal of Strawson’s critic commits the critic to an unnecessarily strong claim, that is, to the claim that reactive attitudes lack propositional content (i.e. they are not about anything) and therefore standard conditions of defensibility and warrant. But, the objection goes, the critic of Strawson need not be committed to this strong a position. He can grant that reactive attitudes have conditions of warrant within the framework of interpersonal relationships, and that these conditions provide a basis for questions about their justifiability. However, the critic will claim that these conditions of justifiability of reactive attitudes are deficient and inadequate in providing a plausible basis for justifying moral criticism.

In adopting such a strategy, some philosophers have purported to identify a normative gap in Strawson’s naturalistic strategy. John Martin Fischer and Mark Ravizza, for instance, argue that:

Strawson’s theory may reasonably be said to give an account of what it is for agents to be held responsible, but there seems to be a difference between being held responsible and actually being responsible. Surely it is possible that one can be held responsible even though one in fact is not responsible, and conversely that one can be responsible even though one is actually not treated as a responsible agent. By understanding responsibility primarily in terms of our actual practices of adopting or not adopting certain attitudes towards agents Strawson’s theory risks blurring the difference between these two issues.36

36 John Martin Fischer and Mark Ravizza, “Introduction,” in their Perspectives on Moral Responsibility
Similarly, in a recent essay Angela Smith has provided a systematic articulation of the gap between holding and being responsible.\textsuperscript{37} Smith’s argument proceeds by distinguishing different senses in which ascriptions of moral responsibility are employed. Smith argues that the locution “$A$ holds $B$ responsible for $X$” (where $A$ and $B$ denote agents and $X$ some action) is open to three interpretations: (1) that $A$ takes $B$ to be open to moral assessment on the basis of $X$ (i.e. $A$ takes $X$ to be a basis for moral appraisal of $B$), (2) that $A$ takes $B$ to be open to moral criticism on the basis of $X$ (i.e. that $A$ takes moral criticism to be an appropriate response to what $B$ has done), and (3) that $A$ blames $B$ (that is, $A$ actually exhibits the attitudes and behaviors associated with blame).\textsuperscript{38}

Smith proceeds to argue that there is an important difference between the first two employments of the locution and the third. According to Smith, the first two employments involve a set of beliefs and judgments about the agent’s responsibility and culpability for his actions. In the case of the first employment, the question is whether the judgment or belief that there exists an appropriate relationship between the agent and his actions (such that the action can serve as a basis for the agent’s moral criticism) is in order.\textsuperscript{39} By the same token, in the second employment, the question is whether the belief or judgment that the agent is culpable for his actions (that is, actually blameworthy) is warranted. This second employment of responsibility presupposes but goes beyond the first employment, since judgments of culpability presuppose that the agent’s actions are an appropriate basis of moral appraisal of the agent, but also hold that the agent’s actions


\textsuperscript{38} Smith, op. cit., pp. 469-70.

\textsuperscript{39} This is what Scanlon calls “responsibility as attributability.” See T. M. Scanlon, \textit{What We Owe to Each Other} (Harvard University Press, 1998) Chapter 6.
flout moral norms.

By contrast, when we use responsibility in the third sense (namely in the sense of active blame), we are concerned not with the question of the appropriateness of judgments or beliefs about the relationship between the agent and his actions. Instead, we are concerned with the question of the relationship that holds between the agent and other moral agents. The question here concerns the implications of the agent’s actions for his relationships with other moral agents. For example, consider a man who drives through city streets at excessive speeds in order to rush his ailing child to hospital. Even if this action is an appropriate basis for moral appraisal of the agent, and even if it is warranted to take the action in question as a basis for morally criticizing him, it is still a further question whether his actions warrant blame, that is, whether his actions impair his relationship with others. This impairment goes beyond the mere judgment (and belief) that the agent is blameworthy, and involve a set of practices and feelings that constitute blame.

What is crucial to note about the dichotomy between on the one hand the first two employments and on the other the third, is that the conditions of appropriateness of the former are different than the conditions of appropriateness of the latter. According to Smith:

...to judge that a person is morally culpable for an action or attitude is to judge that she is responsible for it and that it is morally wrong or unjustifiable, and to actively blame her is to have and perhaps also to express blaming attitudes toward her on the basis of this judgment. The judgment that a person is morally culpable for some attitude thus forms an

40 See, for instance, T. M. Scanlon, *Moral Dimensions* (Harvard University Press, 2008), Chapter 4. Scanlon argues that to blame someone is neither merely an evaluation of the agent, nor a sanction. Instead, blame indicates an impairment of the agent’s relationship with others. In this way, responsibility attaches to the meaning (or significance) an action has, and not to its permissibility.
essential part of the justification for these blaming attitudes and responses. But...the question whether it would be appropriate for any particular person to have and express any particular attitude to the agent will depend upon many considerations in addition to the agent’s responsibility and culpability for the thing in question.\textsuperscript{41}

According to Smith our convictions about when and how it would be appropriate to react negatively to agents (that is, it exhibit the range of attitudes and practices associated with blame) is often influenced by factors which have little or nothing to do with the agent’s culpability for his actions or attitudes.\textsuperscript{42} Smith goes on to enumerate factors which figure into our decision whether or not to engage in a blaming behavior towards someone.\textsuperscript{43}

These include whether the moral judge has the right standing to issue the negative reactions,\textsuperscript{44} the extent of the fault in question,\textsuperscript{45} and the agent’s own attitude towards his fault.\textsuperscript{46} This, of course, may not be an exhaustive inventory. Nevertheless, what is essential to note here is that none of these grounds engage the substantive question of whether the belief or the judgment that the agent \textit{is} an appropriate candidate for moral criticism is defensible. It is, according to Smith, this central feature of blame (as constituted by Strawsonian reactive attitudes), that is, its liability to being influenced by morally extraneous factors, which creates the conceptual gulf between holding responsible and being responsible.

Smith concludes that we must afford \textit{judgments} of agents \textit{being} responsible for

\textsuperscript{41} Smith, “On Being Responsible and Holding Responsible,” p. 477.
\textsuperscript{42} Smith, op. cit., p. 466.
\textsuperscript{43} Smith, op. cit., pp. 478-83.
\textsuperscript{44} One might consider one’s superior rude for being consistently late for meetings, but one may be unwilling to actively blame him, e.g. to show resentment, to remind him of the importance of punctuality, etc.
\textsuperscript{45} Even if one had no reservations about actively blaming one’s superior for being consistently late, one might consider the fault insignificant (either in the context of the agent’s virtues, or else to render appropriate active blame.
\textsuperscript{46} For example, if one knows that one’s superior is already aware of his bad habit and is struggling against it, one might be less willing to engage in active blame.
their conduct conceptual priority over the practices of holding them responsible for their conduct. On Smith’s account, the naturalist interpretation of the quality of will gets the direction wrong. That is to say, it explicates what it is to be an appropriate candidate for moral criticism in terms of the practices of agents holding each other responsible under appropriate conditions. But conditions of appropriateness of blame (as embodied by Strawsonian reactive attitudes) do not necessarily track the moral features of our practices. The naturalist interpretation, therefore, cannot maintain a principled distinction between being responsible and being held responsible.

We can therefore construe both Fischer and Ravizza and Smith to be combating a central feature of Strawsonian compatibilism, namely the claim that there is no appropriate analysis of the notion of being responsible independent of the notion of holding responsible. Thus contrary to the critic confronted earlier in this section, Smith’s criticism (like that of Fischer and Ravizza) does not rest on the claim that reactive attitudes lack propositional content and therefore standard justifiability conditions. Instead, she claims that given what reactive attitudes are (i.e. given that they are natural features of human psychological constitution), there is a discord between conditions of justifiability of reactive attitudes (i.e. blame) and the judgments and beliefs necessary for culpability. According to Smith, this shows that even if the original criticism of Strawson is too crude, our natural propensity to reactive attitudes cannot serve as a basis for a compelling account of conditions of responsible agency.

I think Smith is entirely correct in her identification of the gap between holding and being responsible, as well as in her subsequent rejection of the conceptual primacy of
holding responsible (in her rejection of the dependence of being responsible on holding responsible). However, what I want to argue in the next section is that this argument is incomplete as it stands. For while Smith’s proposal successfully avoids the difficulties I have identified earlier in this chapter in the charge of non-cognitivism against Strawson, she does not advance a positive thesis about the conditions of the appropriateness of the range of judgments constitutive of being responsible. In other words, even granting Smith’s argument that there is a gap between being responsible and holding responsible, and the further claim that such a gap is characterized by the discord between the conditions of justifiability of reactive attitudes and the conditions of justifiability of the range of judgments and beliefs necessary for being responsible, it remains to be seen precisely what conditions are necessary for the appropriateness of the latter, that is, what being responsible consists in, such that our practices of holding responsible fail to do justice to it.

This is the question which I will investigate in the remainder of this chapter, as well as in the subsequent chapters. In Section 2.4 I argue that the fact that, on the prevailing readings of Strawson, the expectations of good will are socially constituted precludes any deeply normative account of conditions of appropriateness of the range of judgments and beliefs indispensable to a plausible account of being responsible.

2.4 Expectations of Good Will and Moral Norms

In Section 2.3 I defended two claims. The first claim was that the prevailing critics of Strawson have relied on an implausibly narrow account of the naturalistic strategy. More specifically, I defended Strawson’s naturalistic strategy against the charge of non-
cognitivism. The second claim was that we require a fuller and more plausible account of the connection between being responsible and holding responsible. If Smith is correct in her criticism of Strawson’s failure to recognize a notion of being responsible independent from our practices of holding responsible, then we do owe an account of what makes the range of judgments and beliefs indispensable to being responsible appropriate.

My aim in this section is to argue that the naturalistic strategy lacks the resources to provide such an account. The difficulty arises, I argue, out of the question of the connection between the naturalistic strategy and the properly constituted moral norms. More specifically, if the expectations of good will are given features of human interpersonal relationships, the range of beliefs and judgments related to being responsible are themselves beyond the reach of a substantive theory of the constitution of moral norms. On my account, the naturalistic strategy is incapable of providing an adequate articulation of the normative basis of expectations of good will, and, in so far as expectations of good will are identified with moral norms, of what justifies moral norms themselves.

To see what the problem is, begin by recalling that according to the quality of will thesis, “Being morally responsible and legitimately holding morally responsible are to be settled exclusively in terms of the moral quality of the will with which an agent acts.”\(^\text{47}\) While I have argued in Section 2.3 that this account successfully blocks the charge of non-cognitivism, it is not clear whether it has the resources to settle the distinct question of whether the naturalistic strategy can isolate the distinctly moral significance of expectations of good will and reactive attitudes that supervene on them. Even granting

that the naturalistic interpretation of the quality of will thesis does not sever the necessary connection between moral criticism and the objects thereof, nothing follows about the moral quality of the will with which the agent acts. In other words, even granting that the quality of will thesis provides a solid basis for ascribing responsibility to the agent, it is not clear what entitles McKenna to further construe such ascriptions as essentially moral in nature.

A similar point can be made against Wallace, who argues that the fact that on Strawson’s account reactive attitudes are essentially connected with expectations of good will “serves to situate moral responsibility in relation to the notion of moral obligation, and helps us see that the stance of holding people responsible involves a commitment to moral justifications, which support the obligations we expect people to comply with.” Thus according to Wallace, the fact that, on Strawson’s account, reactive attitudes supervene on certain expectations, in particular on expectations of good will, makes such attitudes essentially moral, because the expectations in question themselves capture the moral demands we place on agents within the community. But while making reactive attitudes contingent on the degree to which the agent upholds certain expectations incumbent upon him does preserve the right propositional content for moral criticism, in the sense that the said attitudes would be responses towards the agent and what he or she has done, it is not clear whether anything about the moral significance of such reactive attitudes immediately follows. More needs to be said about what entitled Wallace and McKenna to construe such attitudes as morally significant.

In order to make this distinction clear, consider two examples. Consider a public

servant (call this the bad employee example). This person is not neglectful of his duties, nor is he incompetent in his work. He is merely not a good manager of time. He is therefore consistently late for work. Such a person clearly consistently violates certain expectations to which he is held at his work place. Not only does the management have a legitimate claim on all employees in terms of their punctuality, employees in the workplace also have legitimate claims on each other in terms of their conduct at work place. As a result of consistently violating these expectations, the employee in question is bound to face reactive attitudes not only from his employers, but also from his colleagues. He is bound to be the target of certain blame-related responses. He will be held responsible for his transgression of the expectations in question. But it is not clear whether such violations of the expectation of good will, and the reactive attitudes they occasion, indeed amount to a moral failure on the part of the offender. The bad employee might well be open to many varieties of criticism. But being a poor manager of time does not immediately seem to be a moral failure of the agent.

Consider a second example (call it the bad citizen example). Certain communities are governed by strict religious-cultural norms. In certain countries, for example, women are expected to fully cover themselves in public. From a social stand-point, this is a bona fide, socially instituted expectation. In many cases, such expectations play a central role not only in the de facto practices of the community, but are also a salient component of the society’s historical narrative. Those who violate such demands even if not prosecuted by the authorities, are liable to a range social sanctions, including, but not limited to, public scorn, stigmatization, etc.. However, it is clear that some given person’s failure to
observe the norms in question (i.e. her failure to fulfill the expectations in question) does not make her an appropriate target for ascriptions of moral responsibility. We might bring a variety of criticisms to bear on the offending agent. We can criticize her on the basis of her violating the putatively divine instructions, or alternatively for showing contempt for the community’s heritage, and so forth. But none of these seem to amount to a moral assessment of the agent.

These two cases have two features in common. In the first place, in both cases the agent is faced with certain expectations. It is unimportant, for our present purposes, whether the expectations are “reasonable” or not. Rather, what is crucial is that the expectations are properly constituted. This is indeed the case in both examples, as in both the agent is faced with a set of socially constituted demands. In the second place, in both cases the agent’s failure to fulfill these obligations exposes him or her to certain reactive attitudes. These reactive attitudes are not merely personal, though they may be. What is crucial to note, however, is that they (i.e. the reactive attitudes in question) also possess the requisite element of abstraction. The reactive attitudes the bad employee and the bad citizen face are not merely responses individuals within these two communities happen to have to them due to a sense of being personally offended. Instead, the reactive attitudes express a collective resentment toward the agent. In both of these two cases, therefore, the elements of the naturalistic interpretation of the quality of will are satisfied. What I wish to argue, however, is that nevertheless the agents in these two cases are not appropriate candidates for moral criticism. This illustrates the failure of the naturalistic interpretation.
The central reason for this failure of the naturalist interpretation is that it is not clear what the normative significance of a set of socially constituted expectations is, and why these agents’ failures to observe the properly (i.e. socially) constituted demands and expectations betrays a moral failing of the agent. It is therefore not clear why adhering to these expectations is morally incumbent upon the agent, if these expectations are not backed up by some more normatively fundamental demands. For in the absence of such a more normatively fundamental fact, the expectations in questions themselves can turn out to be mere expressions of the community’s preferences, and not connected, in any interesting way, to deeply moral considerations. In the bad employee example, for instance, there seems to be nothing distinctly moral about the punctuality requirement, and therefore the violation of such a requirement does not seem to immediately amount to a moral failure of the agent. We can put the point even more strongly in the case of the bad citizen, where we have reasons to believe that not only are the expectations to which the agent in the example is held are not moral, but that they are decidedly immoral, in so far as they are inconsistent with the agent’s autonomy.

What the proponents of the naturalistic strategy require is an account of the connection between the expectations of good will and the properly constituted moral norms. Such a connection, however, is precluded by a feature inherent in the naturalistic interpretation of the quality of will thesis, namely that on this interpretation, the expectations in question are brute, given facts about human psychological constitution. It is important to emphasize that I am not claiming that such a connection cannot exist if the naturalistic interpretation of the quality of will thesis is true. Consider, for example,
crimes that undermine the victim’s bodily integrity. Such crimes undermine a set of expectations that are given features of our interpersonal interactions. At the same time, these expectations are informed by a set of properly constituted moral norms about the importance of preserving the bodily integrity of persons. In this way, the naturalistic understanding of expectations overlaps with the substantive reasons that inform moral norms. But such a connection, when it does exist, is accidental and not necessary. I take this to be the main moral of the examples of the bad employee and the bad citizen. More specifically, the thought here is that not every properly constituted expectation amounts to a morally significant expectation, that is, an expectation that can give rise to moral criticism. An argument is needed to establish that a set of properly socially constituted expectations is morally significant. However, since the naturalistic strategy takes these expectations to be merely natural facts about human psychological constitution, it precludes any more normatively fundamental account of the norms in question.

I have argued that the naturalistic interpretation of the quality of will thesis is incapable of providing a plausible account of the necessary connection between on the one hand the socially constituted expectations of good will, and on the other bona fide moral norms. But if there is no necessary connection between the expectations of good will (as understood by the naturalistic interpretation of the quality of will thesis) and the moral norms, then the reactive attitudes which supervene on expectations of good will cannot necessarily amount to moral criticism. A similar objection can be advanced against Smith’s distinction between being responsible and holding responsible. More specifically, one can grant that there is a conceptual gulf between on the one hand
judgments of moral responsibility, and on the other active blame (as constituted by Strawsonian reactive attitudes), but this does not settle the question of conditions under which a judgment of moral responsibility is warranted. We can grant Smith’s claim that traditional interpretations of Strawson, with the emphasis they put on active blame, have failed to provide a complete account of conditions of moral criticism. But even then we still seem to lack a compelling account of the distinction between judgments of moral responsibility (as informed by bona fide moral norms) and other non-moral judgments. In this way, emphasizing the conceptual gulf between judgments of moral responsibility and active blame does not provide a compelling account of the conditions of applicability of bona fide moral norms.

In this section I have argued that the naturalistic strategy lacks the resources to offer a principled account of the content of moral norms. I have further argued that Smith’s distinction between being responsible and holding responsible, while allowing us to diminish the importance of reactive attitudes, also lacks the resources to articulate a distinction between judgments of moral responsibility and other judgments.

2.5 Concluding Remarks

My main aim in this chapter has been to offer a critical analysis of Strawson’s naturalistic strategy. In Section 2.2 I argued that Strawsonian non-reductive naturalism aims to neutralize the threat of mechanism by discounting the relevance of metaphysical considerations. This is done, as I have argued, by appealing to natural facts about human beings, viz. the fact that human beings are by nature social and engage in a variety of social transactions, the fact that human beings are liable to holding each other to
expectations which track the agent’s interests, the liability of human beings to a certain class of attitudes and feelings, and the role the natural mechanism of sympathy plays in the development of moral sense. On the Humean interpretation of Strawson, therefore, to ask *why*-questions about our moral practices (that is, to ask for general justification for them) is misguided. The right sort of questions to ask are *how*-questions, that is, questions about how natural faculties with which human beings are endowed explain our actual moral concepts and practices. It is only through tracing our moral practices to our natural human faculties that we can gain insight into the nature of our moral practices and concepts.

In Sections 2.3 and 2.4 I argued that the problem with the naturalistic strategy is not in the absence of conditions of warrant for the application of reactive attitudes, but rather in its inability to offer a plausible account of the normative foundations of expectations of good will. On my account, the problem with Strawson’s naturalistic strategy is not in its inability to offer a plausible and principled distinction between being and holding responsible. The more pressing difficulty is with its inability to articulate a principled distinction between on the one hand being and holding responsible, and on the other being *morally* responsible and holding *morally* responsible.

In the next two chapters, I turn my attention to Strawson’s rationalistic strategy. This strategy proceeds by examining the conditions under which reactive attitudes may appropriately be modified or suspended. In Chapter Three I investigate the question of exemptions, and in particular whether some appropriate analysis of moral capacity is a necessary condition responsible agency. I argue that the function of reactive attitudes, and
the content of moral address, diminish the plausibility of the view that the possession of moral competence is a necessary condition of responsible agency. In Chapter Four I turn my attention to excusing considerations, and in particular to the question of the role of intention in appropriateness of moral criticism. Voluntarist accounts of conditions of moral criticism hold that intention is the primary attitude with which moral criticism is concerned. I argue that cases of omission and negligence diminish the plausibility of this view. In concert, the arguments of Chapters Three and Four illustrate the failure of Strawson’s rationalistic strategy.
CHAPTER 3: EXEMPTIONS AND THE CONDITION OF MORAL SENSE

3.1 Introduction

In Chapter Two I argued that the naturalistic strategy is incapable of offering a compelling basis for the range of attitudes and practices associated with moral criticism. This motivates an interest in Strawson’s rationalistic strategy. The thought here is that if the natural facts about reactive attitudes and expectations of good will marshaled by the naturalistic strategy are incapable of neutralizing the threat of the Causal Thesis, then we ought to seek to locate conditions of responsible agency in a thicker conception of our practices of moral responsibility, and in particular in a compelling account of conditions under which the modification or suspension of reactive attitudes is warranted.

In Section 3.2 I offer a detailed account of Strawson’s rationalistic strategy. The central question with which I will be concerned in that section is the connection between the quality of will thesis and counterconditions to moral criticism, that is, conditions under which moral criticism, as embodied by Strawsonian reactive attitudes, can be justifiably modified or suspended. In the remainder of this chapter I will focus on a particular sub-class of counterconditions to moral criticism, namely exemptions. Exemptions do not deny that the agent has failed to fulfill the demands incumbent upon him, but rather claim that the agent lacks the requisite moral capacities for responsible agency.

Some philosophers writing within the Strawsonian tradition have argued that whether an agent is an appropriate candidate for moral criticism depends crucially on whether he possesses capacities for being susceptible to reactive attitudes. In a recent
essay, Paul Russell has defended such a position. According to Russell, agents who lack the faculty of moral sense (the capacity of feeling and understanding the significance of reactive attitudes for interpersonal relationships) are inappropriate candidates for moral criticism because the absence of the faculty of moral sense entails that such agents are incapable of grasping and applying moral norms. Russell’s defense of the condition of moral sense raises important questions about the relationship between susceptibility to reactive attitudes, the capacity to grasp moral norms, and appropriateness of moral criticism. In this chapter I am interested in two questions in particular. The first question concerns the connection between the failure to be susceptible to reactive attitudes and appropriateness of moral criticism. In Section 3.4 I argue that a plausible account of the function of reactive attitudes diminishes the plausibility of the claim that agents who lack the faculty of moral sense are inappropriate candidates for moral criticism. In Section 3.5 I turn to the broader claim that the capacity to grasp moral norms is a necessary condition of appropriateness of moral criticism. I argue that this claim depends on an account of the content of moral address that is conceptually inadequate. This not only diminishes the plausibility of the condition of moral sense, but also provides a basis for rejecting the initially compelling thought that some notion of moral competence is a necessary component of a plausible account of conditions of responsible agency.

3.2 The Argument from Excuses

As I have argued earlier, on Strawson’s account reactive attitudes play a crucial role in interpersonal relationships, such that living in a world in which there is a thorough objectivity of attitude is not only impossible, but also, even if possible, nevertheless
undesirable. On Strawson’s account, this is because a thorough renunciation of reactive attitudes would have profound implications for how we interact with each other. At the same time, it is clear that it is not the case that such practices are always appropriate. In particular, there are conditions under which we judge that it would be inappropriate to direct towards the agent the full spectrum of reactive attitudes. What are these conditions?

Strawson divides these counterconditions of responsibility into two general groups. To the first group, which Russell terms “specific” considerations, and Wallace “excuses,” belong such considerations as “He did not mean to do so,” or “he did not realize doing so would harm you,” and so forth. Considerations which fall within this category do not invite us to view the agent himself as anything but a fully responsible one, one who is a perfectly apt candidate for reactive attitudes, as well as the range of demands endemic to interpersonal relationships. Instead, excuses suggest that the fact that someone caused a certain injury is perfectly compatible with the fulfillment on his part of the expectation of good will and due regard. The fact of the injury itself is to be viewed as unfortunate and one which is not an appropriate target of moral criticism. In this case, what is being excused is not the agent, but rather the injury, and this invitation to suspend, either temporarily or for a longer period of time, our reactive attitudes towards the agent is done on the basis of the fact that the circumstances in which the

---

injury happened were abnormal.  

Strawson divides the second group of excusing considerations, Russell’s “global” considerations and Wallace’s “exemptions,” into two sub-groups. The first sub-group, consists of considerations which paint the agent as acting out of character (e.g. “He wasn’t himself.”). The appeals of the first sub-group do not invite us not to see the actions of the agent as offensive or contrary to the demands of good will, but rather give us reason to doubt whether the actions of the agent are consistent with his character. For example, when we excuse a grief-stricken friend for failing to uphold the expectations imposed on him by the relationship of friendship, we do so because our antecedent knowledge of him and his character, coupled with the fact that he acted under duress, leads us to suspect that his actions are not an accurate reflection and expression of his considered moral judgments. Given that moral criticism is not only assessment of what the agent has done, but also of himself, of the sort of moral agent he is, holding him accountable for his actions does not seem appropriate.

The second sub-group of exempting considerations, on the other hand, does not give rise to the same kind of concerns. It is not denied by the considerations of the second subgroup that the agent who performed the act, acted in character or that he possessed the requisite kind and degree of control. Rather, such considerations point to the absence of certain capacities in the individual. Exemptions render reactive attitudes inappropriate by

---

4 There is a noteworthy difference between excuses and justifications. Whereas excuses deny that the agent failed to fulfill the basic demand in question, justifications do not necessarily deny this, but rather give reasons why the action in question, even if contrary to the basic demand, was either good or not bad. See J. L. Austin, “A Plea for Excuses,” in *Proceedings of Aristotelian Society* 57 (1957), pp. 1-30; Gary Watson, “Responsibility and the Limits of Evil, Variations on a Strawsonian Theme,” in Ferdinand Schoeman, ed., *Responsibility, Character and the Emotions* (Cambridge University Press, 1987), pp. 256-86, at p. 260.

suggesting that the agent lacks the capacities necessary for being a full participant in normal adult interpersonal relationships, and is therefore an inappropriate candidate for the range of demands and expectations constitutive of interpersonal relationships. Within the second sub-group Strawson includes “he’s only a child.” A very young child typically lacks the range of cognitive and affective faculties necessary for being a full and equal interpersonal relationships. As a result, we tend to modify our expectations of children, as well our reactive attitudes towards them.⁶

Excuses and exemptions give us reason to either modify our reactive attitudes, or withdraw them altogether, either temporarily or permanently, but for different reasons. Excuses block attributability, as they suggest that the injury cannot be attributed to a failure of the agent to uphold the requirements of good will. Exemptions, by contrast, block reactive attitudes directly. Exemptions suggest that even if the agent satisfies the conditions of attributability of some moral failure, his moral incapacity renders him an inappropriate candidate for the range of demands and expectations endemic to interpersonal relationships.

In this chapter I am interested in exempting considerations. Two questions are significant. First, what are the competences that are, on a Strawsonian account, necessary for the appropriateness of expectations and demands endemic to ordinary adult interpersonal relationships? Second, is Strawson correct in claiming that some appropriate analysis of the notion of moral competence is a necessary condition of

⁶At the same time, Strawson realizes that this need not be unqualified. For, as children grow, and become increasingly exposed to, and indeed in many instances integrated into, adult interpersonal relationships, they start to become more and more capable of being full participants in the moral community (p. 75). Indeed, holding them to the same expectations as others is precisely a necessary ingredient of children’s moral education. See John Martin Fischer and Mark Ravizza, Responsibility and Control (Cambridge University Press, 1998) p. 208.
responsible agency. I investigate these two questions in turn.

3.3 **Condition of Moral Sense**

In a recent essay, Paul Russell has defended what he calls the *condition of moral sense*, according to which “the responsible agent must be able to feel and understand moral sentiments or reactive attitudes.”

7 The condition of moral sense, therefore, has both affective (i.e. phenomenological) and cognitive components. It holds that an agent is an appropriate candidate for moral criticism only if he has the cognitive capacity to understand the normative importance of reactive attitudes, and the role they play in enriching human life, as well as the affective capacity to be affected by this understanding. In defending condition of moral sense, Russell seeks to reject the traditional dichotomy between the capacity of responsiveness to moral norms on the one hand and the capacity to be susceptible to reactive attitudes on the other. The traditional accounts hold that the question of whether an agent has the capacity to feel and understand reactive attitudes is distinct from the question of whether he has the capacity to be sensitive to moral norms and reasons. 8 According Russell, by contrast, the capacity to be responsive to moral norms and the capacity to feel and understand reactive attitudes mark the same competency.

---

8 See, for example, John Martin Fischer and Mark Ravizza, *Responsibility and Control*. According to Fischer and Ravizza, an agent is an appropriate candidate for moral criticism if and only if he possesses what they call “guidance control” over the action (p. 34). Guidance control obtains where the agent’s actions issue from his own, reason-responsive deliberative mechanism (p. 39). There are two important points to note about the notion of guidance control. First, Fischer and Ravizza give the requirement of ownership an explicitly subjectivist treatment. They argue that the agent has ownership of his deliberative mechanism if and only if the agent “takes responsibility” for it, which they understand as the agent seeing himself as a causal agent in the world and as an appropriate candidate for the range of reactive attitudes that arise from his agency (pp. 210-212). Furthermore, and this is the second point worth noting, on Fischer and Ravizza’s account the subjectivist-historicist requirement of ownership is logically distinct from the time-slice requirement of reason-responsiveness (pp. 73, 195-196).
Introducing ‘Jack,’ an agent who lacks the faculty of moral sense, Russell writes:

...in an important dimension Jack is not a full participant in (normal) moral life. There is a considerable sphere of moral experience that is simply closed off to Jack. In our dealings with any individual of this kind it is both unreasonable and unfair to communicate and reason with him as if this incapacity was irrelevant to this person’s ability to function as a moral agent.... At the same time, it is important not to exaggerate this problem (significant as it may be). Clearly Jack is an intelligent person and understands moral rules and expectations and the associated sanctions.... Nevertheless, when it comes to the dimension of moral life that involves reactive attitudes, the most that Jack can do is “parrot” these responses or feign feelings of this kind.... in Jack’s world moral considerations lack any emotional coloring of the kind provided by moral sense. For this reason we must conclude that Jack’s moral world-- the way that he experiences it and is moved and directed within it-- is very different from our own.9

On Russell’s analysis, agents who lack the capacity to feel and understand moral emotions are both abnormal and morally incapacitated. They are abnormal in so far as the fact that they are incapable of feeling and understanding reactive attitudes sets them apart from everybody else. I take it that Russell is not of the view that the mere fact of an agent’s abnormality undermines his appropriateness as a candidate for moral criticism.10 Instead, the question we are interested in is what normative capacities are undermined by the inability of the agent to feel and understand reactive attitudes.

According to Russell, agents who are impervious to reactive attitudes are morally incapacitated, in so far as their failure to feel reactive attitudes entails that they fail to appropriately “grasp and apply” moral considerations.11 Russell concedes that to have a diminished capacity to grasp and apply moral norms is not equivalent to being incapable of knowing what moral norms obtain.12 Instead, he claims that such an agent

---

...is [only] able to guide and motivate his conduct with a view to “external” sanctions such as rewards and punishments. Related to this, [he] also has an “external” interest in not arousing negative moral sentiments, since this will obviously affect the way that others treat him. What [he] will lack, nevertheless, is an “internal” system of sanctions (or incentives) as associated with moral sentiments.... For a normal moral agent, our capacity to experience and feel moral sentiments, toward ourselves and others, is intimately and inextricably connected with our understanding of the significance of the background moral claims and considerations.  

On Russell’s account, agents who lack the faculty of moral sense cannot come to “internalize” demands of morality in the way agents with the faculty of moral sense can.  

Agents without the faculty of moral sense may do the right thing, but their actions are not the result of a deep commitment to moral norms, a commitment that only agents who see themselves and others as vulnerable to reactive attitudes can exhibit. Rather, they merely guide their conduct on the basis of the possible rewards and punishments. Intuitively, such agents are missing something crucial about what it is to be a moral agent. Russell attributes this to the role moral education plays in the development of moral sense and the ability to grasp moral norms. A component of moral education is instilling in children the sense that things they do have consequences, and, furthermore, that based on consequences which their agency brings about, they are liable to face reactive attitudes. Therefore, the absence of the faculty of moral sense is symptomatic of the failure of the process of moral education. This must have implications for how the agent understands morality.

We can therefore immediately distinguish two claims explicit in Russell’s defense of condition of moral sense. According to the first claim, agents who lack the faculty of

---

15 Russell, op. cit., p. 296.
16 Fischer and Ravizza, Responsibility and Control, p. 208.
moral sense therefore lack the ability to grasp and apply moral norms, where this is understood as the capacity to act from an internalized concern for the applicable moral norms and the significance of these norms for other agents and his relationship with them. Russell takes this to justify a further claim, namely that agents who lack the faculty of moral sense are inappropriate candidates for moral criticism.

Is it the case, as Russell claims, that agents who lack the capacity to feel and understand reactive attitudes (and are therefore incapable of internalizing moral norms) are inappropriate candidates for moral criticism? I consider this question in the next section.

3.4 Reactive Attitudes, Internalization of Moral Norms and Moral Criticism

Let us first examine the claim that the agent’s ability to internalize moral norms is a necessary condition of appropriateness of moral criticism. In this section I argue that this claim is false. This argument will proceed in two stages. In the first stage I argue that whether an agent is an appropriate candidate for moral criticism is centrally a question about the content of his attitudes, and in particular the reasons which render his attitudes intelligible. I then argue that the question of whether an agent is capable of internalization of moral norms does not bear on the question of whether he is capable of attitudes with the sort of content that is open to criticism in accordance with the quality of will thesis. I conclude that the failure of the agent to internalize moral norms does not render him an inappropriate candidate for moral criticism.

We should begin by clarifying the connection between moral criticism on the one hand and the quality of will thesis on the other. Recall that according to the quality of will
thesis, an agent is an appropriate candidate for moral criticism just in case his attitudes betray a failure on his part to observe the legitimate claims and entitlements of other persons, as codified in moral norms. But whether this is the case depends on the reasons for which the agent performs the action. To see this, let us juxtapose two cases. In one case, you have just defeated me in a contentious game of chess. Out of malevolence and contempt for you and your victory, I step on your foot, which invariably inflicts some pain and injury on you. By contrast, suppose I step on your foot not because of contempt or malevolence, but rather because I see on your foot what I (reasonably) take to be a poisonous spider, and I judge that unless I kill the spider by stepping on it, the spider would bite and seriously injure you.\textsuperscript{17}

There are two points to note about this example. In the first place, moral criticism, as embodied by Strawsonian reactive attitudes (e.g. resentment and indignation) is an inappropriate and unwarranted response in the second case, but not in the first, even if the pain I inflict in the second case is equal to the pain I inflict in the first case. This is the case because my actions in each case are made intelligible by different reasons.\textsuperscript{18} In the first case, my action is made intelligible by my contempt and disregard for you, whereas in the second case my action is made intelligible by my concern for your well-being and comfort. It is this content, that is, the way reasons inform my actions, that reveal something about me and my commitment to moral norms and your claims and entitlements.\textsuperscript{19}

\textsuperscript{18} Scanlon, What We Owe to Each Other, pp. 18-19.
I have argued that, according to the quality of will thesis, whether an agent is an appropriate candidate for moral criticism is centrally a question about the content of his attitudes. I now argue that the question of whether the agent is capable of having attitudes with a content that would engage the quality of will thesis cannot be settled by examining the question of whether he views himself as an appropriate candidate for moral criticism.

In order to illustrate this, consider Alfred Mele’s example of Phil, a philosopher who becomes a hard determinist, and who, as a result, “no longer believes that there [exist] fair targets of the reactive attitudes,” and who does not have “mere residual doubts” that reactive attitudes are metaphysically unjustifiable; he is convinced that they are...”20 Such an agent wholly lacks the ability to internalize moral norms. He may do the right thing, but this is not an indication of a deep commitment on his part to moral norms, or a concern with how his actions impact others and his relationship with them. I want to argue, however, that it is a further question whether he is an inappropriate candidate for moral criticism (that is, whether he is entitled to an exemption).

Consider, for example, the nature and quality of Phil’s friendships. Since Phil is, ex hypothesi, incapable of acting on the basis of the right sort of reason, he fulfills his duties towards his friends not on the basis of a sense of the importance of the relationship of friendship, or out of respect for the unique demands his friends have on him in virtue of standing to him in this particular relationship. Instead, he does so on the basis of a calculus of the benefits the relationship of friendship would bestow upon him.21 Phil

21 Scanlon, What We Owe to Each Other, pp. 159-63.
judges that maintaining a large circle of friends would not only elevate his social status, but would also be useful in times of financial or emotional need. We are likely to find that the quality of our interactions with Phil is severely diminished, perhaps to the extent that the severing of the relationship is warranted. But note that it is a further question whether such a person is an appropriate candidate for moral criticism. The fact that the agent is motivated by wrong reasons may change how we interact with him, but not whether he is accountable for what he does. This is evident from the betrayal we would feel if we find out that someone we considered a friend honored his friendship obligations for reasons of the wrong sort. We would not react to such a person by excusing his actions. Instead, resentment, indignation, censure and ostracism seem appropriate.

This suggests that Russell fails to adequately distinguish two different questions. One is the question of what moves the agent to action. This is a question about the agent’s motives and reasons for acting. It is a logically distinct question, however, whether the agent is an appropriate candidate for moral criticism. While agents who lack the capacity to feel and understand reactive attitudes as a result lack the capacity to act for the right sort of reasons, and with a full and robust understanding of the range of moral claims and their normative background and significance, it is an entirely different question whether this is an appropriate basis for an exemption. For the inability of an agent to feel and understand reactive attitudes to constitute an exemption, it must be the case that agents who are thus incapacitated are therefore incapable of having a morally assessable quality of will. In other words, to sustain the condition of moral sense, we require a conceptual link between reactive attitudes on the one hand and the quality of will thesis on the other.
This connection is precluded, I argue, by the different roles reactive attitudes and the quality of will thesis play. Reactive attitudes play a dual expressive role. In the first place, they give public form to expectations we have of each other. Reactive attitudes are a medium for the communication of what we expect from each other. Examples range from mundane to morally salient. If I consider it inappropriate for a colleague to borrow my books without permission, I can convey this through reactive attitudes, by displaying various reactive attitudes should these expectations go unfulfilled. The case of children’s moral education also provides a good illustration of this function of reactive attitudes. Children are not born with a full understanding of moral norms. At least part of instilling in them an understanding of the difference between right and wrong, and enabling them to navigate the complicated maze of every day interpersonal transactions, is illustrating the connection between reactive attitudes and transgression of moral norms. This establishes, in the mind of the child, a constant conjunction between reactive attitudes on the one hand and transgression of (moral) norms on the other, giving rise to the idea that reactive attitudes are a good indication of what expectations obtain.

The second expressive function of reactive attitudes is to convey the consequences and implications of the agent’s failure to observe the relevant expectations. Reactive attitudes have a communicative power that mere words lack. For example, suppose a colleague plagiarizes one of my essays. In such a case, the agent has failed to uphold certain morally salient expectations. Reactive attitudes function to convey the significance this transgression has for me, and for my relationship with the offender. For example, if I begin resenting the colleague in question and being indignant towards him,
or restrict our interactions to a simple passing ‘hello’ in the hallway, this would be
effective in communicating to my colleague the implications his transgression has for me
and my relationship with him. In such a case, my reactive attitudes are symptomatic of
the impairment of my relationship with the offending agent.

But what is crucial to note is that while reactive attitudes serve these expressive
functions, they do not \textit{constitute} the moral norms that obtain.\footnote{Cf. Jesse Prinz, “On the Emotional Bases of Moral Judgment,” in \textit{Philosophical Explorations} 9 (2006) pp. 29-43.} In other words, the reason
that my colleague has wronged me by plagiarizing my essay is not that his plagiarizing
my essay has evoked, or tends to evoke, a particular response from me. There are
independent reasons that legislate against theft of ideas. My reactive attitudes merely put
these reasons in a different light, that is, they endow these reasons with a particular
significance for me and my relationship with others. Reactive attitudes, therefore, do not
constitute what it is to wrong or benefit another. The important point here is that the
question of whether the agent is one who is capable of having attitudes of the requisite
content, that is, attitudes that are informed by reason and therefore can be subject to the
quality of will thesis, is logically independent of the question of whether the agent is
capable of \textit{seeing} himself as such. Even agents who lack the ability to internalize moral
norms are capable of failing to uphold the requirements of good will. This is because the
fact that they are not susceptible to reactive attitudes does not, by Russell’s own
admission, entail that they are incapable of acting for reasons (it just means that they tend
to act for the wrong reasons).

In this section I have argued that the fact that an agent lacks the capacity to
internalize moral norms does not entail that he is an inappropriate candidate for moral
criticism. This supplies one reason for rejecting the condition of moral sense. In the next
section I argue that even granting that agents who are incapable of feeling and
understanding reactive attitudes are therefore incapable of grasping and applying moral
norms, it does not follow that such agents are inappropriate candidates for moral
criticism. This is because, contrary to the prevailing view, the capacity to understand
moral norms is not a necessary condition of appropriateness of moral criticism.

3.5 Understanding Moral Norms as a Precondition of Moral Criticism

Traditionally, the capacity to understand what moral norms obtain has been taken to be a
necessary condition of appropriateness of moral criticism. R. Jay Wallace, for example,
has argued that this constraint on moral criticism arises out of considerations of fairness.
According to Wallace:

[Conditions of accountability] must make it fair to demand that [some
agent] s comply with moral obligations we accept, where the obligations
are supported by distinctively moral reasons. But it would seem fair to
demand this of a person only if the person possesses what I shall refer to as
the powers of reflective self-control: (1) the power to grasp and apply
moral reasons, and (2) the power to control or regulate his behavior by
the light of such reasons. 23

Wallace’s argument consists of two claims. The first claim is that the question of whether
an agent is an appropriate candidate for moral criticism (that is, the question of what
warrants the range of practices and attitudes constitutive of moral criticism) is governed
by norms of fairness. 24 On Wallace’s account, it is the notion of fairness that bridges the
gap between holding responsible (as constituted by reactive attitudes) and being

responsible. The second claim is that moral criticism is warranted and appropriate only if the agents possess what Wallace calls the power of reflective self-control, that is, the normative competence to understand moral norms, as well as the practical capacity to govern their conduct in light of their understanding of what morality requires.

Wallace’s requirement of reflective self-control gives rise to two important questions. The first question is whether Wallace is correct that appropriateness of moral criticism is governed by norms of fairness. The second question is concerned with the broader issue of the relationship between on the one hand the justifiability of moral criticism and on the other the notion of moral competence. Is Wallace correct in his claim that moral criticism is rendered unwarranted and inappropriate by the absence of normative competence? In what follows I will not be interested in the requirement of fairness, that is, the question of whether moral criticism is governed by norms of fairness. Instead I will be interested in the question of whether moral criticism is rendered inappropriate by the absence of moral competence.

One rationale offered in support of the view that moral competence is a necessary condition of appropriateness of moral criticism concerns the requirement of intelligibility of the demands implicit in moral criticism. Gary Watson, for instance, has highlighted the importance of the notion of moral address to Strawson’s account of reactive attitudes.\(^{25}\) According to Watson, implicit in moral criticism (in such attitudes as resentment and indignation, as well as such practices as admonishment, censure and isolation) is a demand for good will.\(^{26}\) Similarly, according to Lawrence Stern reactive attitudes

\(^{25}\) Watson, “Responsibility and the Limits of Evil.”
\(^{26}\) Watson, op. cit., pp. 264-5.
“express our demand that persons extend a modicum of good will not only to ourselves but also to others.”

In this way, moral criticism (and reactive attitudes in particular) supply what Stephen Darwall calls second-personal reasons, that is, reasons whose validity depends on the relationships that hold between the addresser and the addressee. However, in order for this demand to be intelligible, that is, in order for it to be successfully conveyed, a number of “felicity conditions” must be satisfied. In particular, the right authority relationship must hold between the agent making the demand and the agent to whom the demand is addressed, and the demand must be addressed under the right circumstances. What is important for my purposes here, however, is that a necessary condition for the intelligibility of the demand implicit in moral criticism is that the agent to whom it is addressed should be capable of understanding the nature of the demand. As David Shoemaker argues:

...if the practice of holding responsible simply consists in the expression of the basic demand [for good will] via the reactive attitudes, then engaging in the practice is intelligible only under the assumption that the target is capable of understanding and responding to such demands. The expression of resentment, after all, reflects “an expectation of, and demand for, the manifestation of a certain degree of good will and regard on the part of other human beings towards ourselves.” Such expectation is simply unintelligible, though, unless the entity one resents is capable of understanding what goodwill is and requires and is also able to transform that understanding into an actual manifestation of it. The type of

---

29 For example, if a friend of mine asks for a small monetary loan, which I can clearly afford, and I nevertheless deny him, it would be reasonable for him to feel resentful. By contrast, if a total stranger made a similar request and I denied it, it would not be clearly reasonable for him to be resentful. What is important here is that the stranger lacks the standing to make the demand in question. See Darwall, op. cit., p. 60.
30 For example, I may well be in a position to object to someone who smokes in a restaurant, but this demand would be inappropriate in an area clearly marked off as permitting smoking.
communication that is constitutive of the expressed reactive attitudes will be pointless as a form of moral address when directed at those without these epistemic and motivational capacities.\footnote{David Shoemaker, “Moral Address, Moral Responsibility, and the Boundaries of the Moral Community,” in \textit{Ethics} 118 (2007) pp. 70-108, at pp. 74-5.}

On this account, to attempt to morally address (i.e. morally criticize) an agent who is incapable of understanding moral norms is tantamount to attempting to carry on a conversation with someone who cannot understanding the meaning of basic units of the language. In both cases, our address is unintelligible. Agents who are incapable of understanding the nature of the demand are not, following the Strawsonian taxonomy, full members of the moral community, and as such are inappropriate candidates for moral criticism.\footnote{Stern, “Freedom, Blame and Moral Community,” pp. 78-9.}

In what follows I argue that this account is false, and that the intelligibility of moral address is not undermined by the inability of the agent to grasp moral norms. In order to make this claim intuitively plausible, I would like to begin by considering Gary Watson’s discussion of the case of Robert Harris. Harris brutally murdered two teenage boys, stole their lunch, and later mused about impersonating law enforcement and witnessing the reactions of his victims’ parents as he informed them about their children’s deaths.\footnote{Watson, “Responsibility and the Limits of Evil,” pp. 268-71.} As Watson notes, such cases pose challenges for our understanding of preconditions of moral criticism. On the one hand, Harris’ extreme brutality, and the blatant disregard he demonstrates for the relevant claims and demands for due regard others have against him suggests that he is an archetypal candidate for the range of practices constitutive of moral criticism. Resentment, indignation and censure are \textit{prima}
appropriate responses to what he has done. Furthermore, his extreme brutality and the calculated manner in which he carries out his crime, and the sophisticated measures he takes to evade capture, suggest that he is not out of touch with reality. At the same time, Harris’ extreme and inordinate brutality gives pause to our reactions to Harris. The problem here is not merely that Harris’ extremely harsh formative circumstances put pressure on our intuitions. The problem goes deeper, namely that Harris’ inordinate brutality is symptomatic of a set of incompetencies that bear directly on his culpability.

Much literature is dedicated to articulating a compelling account of the incapacities that underlie this intuition. Empirical and philosophical studies have established that psychopathy is characterized by emotional abnormalities such as imperviousness to fear, and lack of empathy. As a result, the psychopath is incapable of sharing the concerns of other persons, that is, incapable of recognizing moral norms as having a basis in the claims and entitlements other persons have against him. This leads Watson to argue that cases such as that of Harris expose an ominous tension in our

---

understanding of conditions of responsible agency. According to Watson:

...if reactive attitudes were implicitly “invitations to dialogue,”...then Harris would be an inappropriate object of such attitudes. For he is hardly a moral interlocutor.... In this instance, an invitation to dialogue would be met with an icy silence (“he has nothing more to say”) or murderous contempt.”

This seems to give rise to a genuine puzzle. If the capacity to understand what moral norms obtain is a necessary condition of appropriateness of moral criticism, then even the most morally depraved agents would be immune from moral criticism in virtue of their moral depravity, a puzzling conclusion for a plausible account of preconditions of moral criticism to reach.

The appearance of a puzzle, however, is misleading. While I think Watson is correct in noting the complexity of our reactions to Harris, his analysis of this complexity is not compelling. While Harris may lack the capacity to grasp and apply moral norms, the atrocities he commits, and his behavior subsequently, constitute compelling evidence of a faulty quality of will. Recall that, as I argued in Section 3.4, whether an agent is an appropriate candidate for moral criticism is centrally a question about the content of his attitudes, that is, about the reasons for which he acts. The same point applies to the quality of will thesis, and in particular to the question of whether an agent, in acting a certain way, fails to show appropriate regard for the legitimate claims of others. Whether an agent shows appropriate regard for the well-being and legitimate claims of others is centrally a question about the content of his attitudes.

Contrast the case of Harris with the case of an agent (Paris) who commits comparably violent crimes, and displays similar lack of remorse. In other words, not only

are the crimes of Paris as brutal, but, just like Harris, he fails to take ownership of his crimes and their consequences. Furthermore, while Paris lacks the faculty of moral sense, he manages to conform with the relevant norms, perhaps out of the fear of punishment, or an extra-ordinary effort of will. Instead, he commits his crimes while sleep-walking, which is the result of a proven neurological disorder. At least intuitively, there is an important difference between these two cases. Moral criticism of Paris is inappropriate and unwarranted, and our intuitions about this are unequivocal (or, at any rate, less equivocal than our intuitions about Harris’ case).

However, this difference cannot consist in either the second agent’s lack of moral sense, nor in the fact that his behavior can be causally explained. This is because, ex hypothesi, these two features are shared by both Harris and Paris. What is crucial about the case of Paris is that the fact that he commits the acts in question while sleep-walking suggest that his actions do not reveal anything about the content of his attitudes. It is this difference about the content of the second agent’s attitudes that informs our intuitions about his culpability.

Having established that the requirement of appropriate regard is centrally a requirement about the content of the agent’s attitudes, the next question is whether the fact that Harris lacks the normative capacities identified by Wallace and Russell has any bearing on the content of his attitudes. Does the fact that Harris lack the capacity to grasp and apply moral norms render his actions altogether devoid of the requisite content that could justifiably attract moral criticism?

The answer to this question is negative. This is because the very fact of Harris’
brutality is revealing about the content of his attitudes, and in particular about the lack of concern he displays for the legitimate claims and well-being of others to whom he owes such duties. He may be performing the actions he does for fun, or to satisfy his need for excitement, etc., but these very features of his character are revealing about the content of his attitudes. They reveal Harris to be someone for whom the well-being of others enjoys little significance, someone for whom the need for excitement trumps concern for the rights of others, etc.. He is indifferent to the suffering of others, and incapable of interacting with them as agents entitled to care and concern. These very features of Harris are themselves legitimate grounds for moral criticism. In other words, the fact of his lack of the faculty of moral sense illuminates, and not distort, the kind of person he is.

It is important to be clear about what this argument commits me to. One way to show the requirement of normative competence to be false is by denying the subordinate claim that implicit in moral criticism is moral address, denying the claim that implicit in moral criticism is a demand that the agent show appropriate regard for the legitimate claims of others. This is not the strategy I have adopted. Instead, I grant the relevance of moral address, but argue that the fact that we have good reasons to take Harris to be an appropriate candidate for moral criticism in spite of his lack of moral sense, suggests that the content of moral address, and therefore conditions of its intelligibility have been misunderstood. Once we arrive at a plausible account of the content of moral address, it will follow that the conditions of its intelligibility, and therefore of appropriateness of moral criticism, are not violated by the absence of the faculty of moral sense.

What is this prevailing view about the content of moral address? I argue that the
offending assumption is that the demand implicit in moral criticism is a moralized demand. Traditionally moral address has been taken to express a demand for reasonable regard. According to Lawrence Stern, for example, reactive attitudes “express our demand that persons extend a modicum of good will not only to ourselves but also to others.”40 Similarly, Watson claims that “negative reactive attitudes express a moral demand: a demand for reasonable regard.”41 The thought here is that when we morally criticize someone, this moral criticism takes place against a background of accepted moral notions and distinctions. The demand implicit in moral criticism, then, is a demand for compliance with these accepted moral norms. Thus the prevailing account takes the relevant moral notions be built into moral criticism, and the moral address implicit therein.

Is it the case that the demand implicit in moral criticism, the demand to “extend a modicum of good” is essentially a moral demand, a demand the presumes a range of moral distinctions, without the capacity to understand which moral address and therefore moral criticism is unintelligible and therefore unjustifiable? In what follows I argue that this understanding of the content of moral address is false.

Recall that I argued in Chapter Two that a naturalist interpretation of the quality of will thesis is incapable of providing an adequate basis for the range of attitudes and practices constitutive of moral criticism. I argued that it leaves mysterious the nature and content of the moral norms which form the bases of moral criticism. The same objection applies to the assumption under consideration, that is, the view that the content of the

demand implicit in moral criticism is moralized. The view in question overlooks the fact that moral criticism is primarily a normative exercise, and, more specifically, an exercise concerned with reasons. Moral criticism claims that the agent has violated a moral norm. Norms express the permissibility or impermissibility of some given attitude or action in a particular set of circumstances. However, norms in general, and moral norms in particular, are not free-standing, but are informed by reasons. There are reasons that count in favor of certain norms, and therefore make them intelligible. For example, moral norms against sexual assault are grounded in reasons about the importance of persons’ autonomy and bodily integrity. To hold that moral address is moralized (i.e. it takes place against a background of accepted and presumed moral distinctions) leaves entirely mysterious the grounds for moral criticism, and in particular the normative underpinnings of the reasonable regard demanded.

This criticism suggest that what is missing from the prevailing accounts of the content of moral address (according to which moral address expresses a moralized demand) is a normatively significant connection between on the one hand the demand expressed in moral address and on the other the entitlements agents have against each other. I argue that this connection can be supplied by the notion of justification.42 Consider the case of a child who is abused by his father, such that when he grows up he develops severe emotional and psychological difficulties. It seems plausible to understand the child’s blame towards his father as a demand for justification, that is, as saying “why did you do it?”. What the victim is demanding of his father is that he (i.e. the father)

either justify his conduct, or else concede that his conduct is not justifiable and proceed to make amends. In doing so, the father recognizes the status of the victim as a being to whom he owes justification.\footnote{Scanlon, \textit{What We Owe to Each Other}, p. 162.} In other words, the sort of demand implicit in the victim’s moral criticism is that he (i.e. the father) should be prepared to justify his past conduct. This characterization is substantiated by the fact that the moral outrage of the victim is compounded should his father either altogether fail to acknowledge his transgression, or, worse yet, condone it, even if the abuse has already ceased. The familiar phenomenon of “need for closure” often encountered by victims of serious crimes reflects not only the inescapability of the need for dialogue, but is also instructive concerning the content of the demand implicit in moral criticism.

I have argued that to morally address someone is to demand that the target of the address engage in a process of justification of his conduct in light of reasons. This demand evolves into moral criticism just in case the purported justification is not satisfactory, or if the agent refuses to acknowledge the demand for justification. There are two features of justification that are important to highlight. First, the demand for justification is not itself a moralized concept. To demand justification does not necessarily presume any moral notions. For example, there is no moral concept presumed in my demand that you justify your partiality towards classical music. There may be reasons that inform your views, but there does not seem to be anything necessarily moral about them.

The second point worth emphasizing concerns the question of conditions under which a demand for justification unintelligible. Granting that the demand implicit in...
moral criticism is a demand for justification, is it the case, as the traditional understanding of conditions of moral criticism holds, that agents who are incapable of understanding moral norms are inappropriate candidates for moral criticism in virtue of the unintelligibility of the demand implicit in moral address?

What is important to note is that whether the demand for justification implicit in moral address is intelligible is not a function of the psychological competence of the agent, but rather a function of the content of the demand for justification. For example, it is normally entirely inappropriate to talk on one’s cell-phone in a theater. In doing so, one simply fails to take note of the salient feature of one’s circumstances. Such a person not only betrays a failure to appreciate the value of the activity he is engaged in, he, more importantly, fails to show appropriate regard for the claims and entitlements of those around him. Such a person displays a faulty attitude, an attitude that betrays a failure to care for the justifiability of his conduct to others. By contrast, sneezing during a play is not typically taken to indicate a faulty attitude on the part of the agent, even if it is just as distracting as using a cell-phone. This is centrally due to the fact that sneezing lacks the requisite content, that is, is not amenable to influence by reasons and therefore impervious to rational justification. While it makes sense to demand justification from someone for his talking on the cell-phone, demanding justification for a sneeze is unintelligible.

What these considerations suggest is that the capacity to grasp moral norms is not a condition of intelligibility of moral address. Thus even granting that the intelligibility of moral address implicit in moral criticism is a necessary condition of appropriateness of
moral criticism, it does not follow that the possession of the capacity to grasp and apply moral norms is a necessary condition of appropriateness of moral criticism.

It may be objected that this conclusion is implausible. This implausibility is best illustrated in the case of children. Clearly, children enjoy a unique moral status. While their actions may inflict harm on others, they seem to be inappropriate candidates for the range of attitudes and practices associated with moral criticism. Resentment, indignation, and censure are inappropriate responses to children’s wrong-doing. It is in recognition of such a fact that Strawson argues that being a child is an among excusing considerations. Any plausible theory of conditions of responsible agency must account for this phenomenon. Therefore, my conclusion to the contrary is implausible and should be dismissed.

But the issue of children’s moral agency, far from discrediting my conclusion, strengthens it. We can begin by noting that children, far from being unequivocally exempt from moral criticism, represent boarder-line cases. As Strawson himself notes, children are incrementally held to be answerable for their faults and mistakes. This inauguration into responsible agency happens through a process of moral education. moral educators (teachers, parents, figures of authority, etc.), instill in them, in the first place, the notion that things they do often have consequences, negative or positive, (“See what you have done,” “you have been very helpful,” etc.), and in the second place, the sense that based on consequences which their agency brings about, they are liable to being praised or blamed (“why don’t you go to your room and think about what you have done,” “well

---

44 See, for example, Strawson, “Freedom and Resentment,” p. 66.
Given this (admittedly brief) account of moral education, I want to insist that the proposition that children are exempted from moral criticism is shown to be false by the very possibility of moral education (and therefore moral agency). If it were true that children were inappropriate candidates for moral criticism, it would be difficult to make sense of the efficaciousness of the process of moral education (and the eventual emergence of responsible agents). In order to see this, it is important to recall that, as I argued in the previous chapter, what distinguishes reactive attitudes from the objective attitude is that reactive attitudes have a felt quality. When an agent fails to fulfill the demands incumbent upon him, this gives rise not only to the one who is wronged, but also to those witnessing the wrong, a series of emotions and feelings of indignation and resentment. What is crucial to note is that it is precisely this felt quality of moral criticism (as embodied by Strawsonian reactive attitudes) that is necessary for the efficaciousness of the process of moral education. This is because a necessary condition of the efficaciousness of moral education is the gradual communication to the agent of the significance and the impact of his actions on others. This is often done not through rational argumentation with the child, but rather through emotional appeals. For example, if a child takes another child’s toys without permission, we might attempt to illustrate the wrongness of this act by asking how he himself would have felt if some does the same to him, by showing disappointment in him, etc..

The expression of these emotions communicate to the child the consequences of

---

his actions, and, relatedly, the significance of his actions for his relationship with others.
But, and this is the crucial point, if reactive attitudes were inappropriate responses to this
infraction, it would be difficult to see how the process of moral education would proceed
at all. It would be difficult to see how one could communicate to the child the information
necessary for the formation of a full and robust sense of moral right and wrong. Thus
unless we are prepared to reject altogether the possibility of moral education (and, by
implication, the possibility of moral agency), we must concede that children are
appropriate candidates for moral criticism, with the proviso that the kind and degree of
moral criticism that is warranted in the case of children are different from those
warranted in the case of adults, as a result of the fact that the aim of criticizing children is
to effect a process of moral education (i.e. give rise to a stable moral character) whereas
in the case of adults, this moral character has presumably already in place. While we may
be indignant and resentful towards an adult who fails to fulfill the relevant demands, this
may transform to mere ‘disappointment’ in the case of children.46

The question I investigated in this section is whether the capacity to grasp and
apply moral norms is a precondition of moral criticism. I have argued for the conclusion
that it is not. The argument proceeded in two stages. First, I argued that the sort of
demand implicit in moral address is not a demand for good will, but rather a demand for
justification. I then argued that whether this demand is intelligible is a function not of the
competence of the agent to whom it is addressed, but rather depends on the content of the
demand for justification, that is, that for which justification is demanded. From these two

46 What is important to note is that disappointment is itself a reactive attitude and possesses a felt quality.
The thought here is that if lack of moral capacity renders moral criticism inappropriate, then even
disappointment would be inappropriate, in which case the very possibility of moral education and moral
agency is threatened.

86
premises follows that it is not the case that agents who lack the capacity to grasp and apply moral norms are inappropriate candidates for moral criticism on grounds of the unintelligibility of the demand addressed to them. Such agents can meaningfully be required to justify their conduct in light of reasons. It is a further question what implications moral criticism ought to have for our relationship with such agent.

3.6 Concluding Remarks

In this chapter I have investigated the question of whether susceptibility to reactive attitudes is a necessary condition of appropriateness of moral criticism. According to Russell’s condition of moral sense, agents who lack the faculty of moral sense therefore lack the ability to internalize moral norms, and this makes them an inappropriate candidate for moral criticism.

In Section 3.4 I argued that a plausible account of the function of reactive attitudes entails that moral criticism is not inappropriate in the case of agents who lack the capacity to internalize moral norms. This is because reactive attitudes play only an expressive function; they do not constitute what wronging another consists in. In Section 3.5 I turned to the broader claim that moral competence is a necessary condition of responsible agency. On such a view, implicit in moral criticism is moral address, and the intelligibility of moral address requires understanding of moral norms by the agent. I argued that this view depends on the assumption that the content of moral address is essentially moralized, but that this assumption is false. I argued that moral address is constituted by a demand of justification, that is, a demand that the agent justify his conduct in light of reasons. Not only is this demand not necessarily moralized, but,
furthermore, the standards of intelligibility of this demand do not depend on the agent’s normative competence, but rather on the content of address, that is, on that for which justification is being demanded. Therefore, even if an agent does lack the capacity to understand moral norms, it does not follow that the demand implicit in moral address is unintelligible.

The falsity of the condition of moral sense has important implications for our understanding of conditions of responsible agency. Even if Russell is correct in claiming that agents who lack the faculty of moral sense therefore lack the capacity to grasp and apply moral norms, it does not follow that such agents are inappropriate candidates for moral criticism. Furthermore, we must turn our attention away from the psychology of responsible agency, and focus on the content of moral criticism and its significance for interpersonal relationships.

In Chapter Four I will focus on the question of the connection between the quality of will thesis and intention, with the aim of clarifying whether an agent can be appropriately morally criticized only if his actions were intended, or can be traced to an earlier intention. In Chapter Five I will explore the question of the connection between moral criticism and reasons. The question that will be of central importance in Chapter Five is whether we can articulate a compelling account of the normative bases of responsible agency that does not invoke considerations about freedom and control.
CHAPTER 4: EXCUSES AND THE REQUIREMENT OF INTENTION

4.1 Introduction

Thus far I have articulated and criticized two competing interpretations of the quality of will thesis. In Chapter Two I focused on the naturalist interpretation of the quality of will thesis, and argued that the natural facts about human psychological constitution on which it is founded are incapable of providing an adequate basis for the expectations of good will and demands endemic to interpersonal relationships. In Chapter Three I turned my attention to Strawson’s rationalistic strategy and his account of exempting considerations. I argued that we not only have good reasons to reject the view that possession of the faculty of moral sense is a necessary condition of appropriateness of moral criticism, but we must also reject the initially appealing thought that the possession of moral competence is a necessary condition of responsible agency.

In this chapter I turn my attention to conditions for the applicability of excuses. I will focus on the voluntarist interpretation of the quality of will thesis. Voluntarist accounts hold that the presence of choice or intention is a necessary condition of appropriateness of moral criticism. The voluntarist interpretation of the quality of will thesis offers a particular rationale for this connection, namely that the absence of an intention on the part of the agent blocks attributability of an action to him. Therefore, in order to show the voluntarist interpretation of the quality of will thesis to be false (and that the absence of an intention does not block attributability), it would not be sufficient to show that moral criticism remains appropriate even in the absence of an intention, though this would be the right general strategy. What is further required is an
investigation into the question of the connection between the requirement of intention and the quality of will thesis.

My central argument in this chapter is that cases of negligence and omission expose a conceptual gap between the quality of will thesis and the requirement of intention. This argument will proceed in two stages. In Section 4.3 I argue that the requirement of intention is centrally a causal requirement. I then proceed to argue, in Sections 4.4 and 4.5, that this feature of intention renders the voluntarist accounts incapable of fully capturing our concern with the content of the agent’s attitudes. I argue that this gulf is illustrated by cases of negligence and omission. In such cases, the very fact that an agent is oblivious to certain features of his situation is sufficient to warrant moral criticism. The thought here is that cases of negligence and omission reveal that intention is not the primary attitude with which moral criticism is concerned.

4.2 Voluntarist Interpretation of the Quality of Will Thesis

Under what conditions can we conclude than a certain harm is not attributable to the agent? In formulating the quality of will thesis, Strawson writes:

> The central common place that I want to insist on is the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions.¹

Strawson goes on to argue that excuses

---

It might therefore seem that on a plausible interpretation of the quality of will thesis, intention is a necessary feature of acts for which agents can be morally criticized. According to this interpretation, excuses and exemptions function by showing that the agent either did not act intentionally, or was incapable of forming morally significant intentions altogether.

To illustrate the importance of intention, consider the speeding driver example, where an agent is driving at excessive speeds, and violates a number of traffic laws and therefore puts the safety and well-being of others in jeopardy. But suppose in one case he is doing this out of a depraved disregard for the well-being of others, whereas in the other case he is doing this in order to rush a member of his family to the hospital. As I have noted earlier, intuitively there is an important difference between these two cases. Whereas resentment, indignation and censure seem appropriate reactions to the first agent, they seem to be inappropriate responses to what the second agent has done. Once we know about the circumstances of the second agent, our reactions to him are bound to be modified. We are bound to either suspend our resentment towards him, or, at the very least, moderate these reactions. By contrast, no further explanation about the first driver’s actions seem to be capable of dislodging the sense of resentment and anger we feel towards him.

A similar conclusion is prompted by the following case, related by Daniel Dennett:

I read in the newspaper recently about a young father who forgot to drop off his infant daughter at the day-care center on his way to work. She spent the day locked in his car in a hot parking lot, and in the evening on his way home when he stopped at the day-care center to pick her up, he was told, “You didn’t drop her off today.” He rushed out to his car to find her still strapped into her little car seat in the back, dead.3

It may be thought that whatever our reaction to this father may be, it cannot be characterized as moral criticism. Resentment and indignation are inappropriate reactions to the scenario. It may of course, be objected that one reason we seem unwilling to blame the father is the vulnerability we all have to being the victim of the circumstances, as the father was. It is entirely conceivable that any one of us would be susceptible to forgetfulness. This moderates our reaction to the father. But even if this is part of the reason for our reaction to the father, it cannot be the whole of it. For even if we somehow had assurances that we are immune to the sort of omission displayed by the father, it would still seem unreasonable to direct at him the full range of reactive attitudes constitutive of moral criticism.

In concert, the case of the forgetful father and the speeding driver motivate a voluntarist interpretation of the quality of will thesis. R. Jay Wallace, for example, has argued that whether or not an agent has violated the requirement of good will is primarily a question about the intention that is reflected in his attitudes. Wallace invites us to consider the duty of nonmaleficence, and argues that:

...this is not simply an obligation not to make bodily movements that harm other people. Rather it is an obligation not to act in ways that express

the choice to harm other people, in the ordinary pursuit of one’s own ends. Accordingly, the primary target of moral assessment in terms of this obligation is not bodily movement per se, nor is it emotions and desires to which we are subject; rather, it is the quality of choice expressed in what we do. Thus we are not blamed for violating the duty of nonmaleficence except when what we do results from a choice to harm someone.4

The central idea motivating Wallace’s voluntarist interpretation of Strawson is that in the absence of the relevant intention, it would be mistaken to say that the agent acted from a bad quality of will.5 If we are to take as the locus of moral criticism the degree to which the agent’s actions betray a morally faulty quality of will, it is only the presence of an intention, which informs the action in question, that can serve as the basis for this moral criticism. On this account, intention is the primary attitude with which moral criticism is concerned.

This is the claim I will investigate in the remainder of this chapter.

4.3 Intention, Control, and Causation

To act intentionally is to exercise control over what one does. The central example of control is what we can call physical control. Daniel Dennett, for example, has characterized control as follows: “A controls B if and only if the relation between A and B is such that A can drive B into whichever of B’s normal range of states A wants B to be in.”6 Dennett then goes on to argue that this conception of control requires that the controller be an entity capable of being in different states, and the controller one among whose psychological repertoire are desires and wants. Dana Nelkin has likewise argued that the concept of free agency (the kind of agency presupposed by our practices of moral

5 Wallace, op. Cit., p. 126.
6 Daniel Dennett, Elbow Room (MIT Press, 1984) p. 52. emphasis in the original.
responsibility) implicit in ascriptions of responsibility is that of one’s actions being *up to one* in such a way that one can be accountable and responsible for them. It is critical to note that the sort of up-to-us-ness that is implicated, according to both Dennett and Nelkin, in our concept of free agency, is essentially a *causal* notion.

Our interest in control has two distinct but related sources. In the first place, the notion of control demarcates the boundary between things that an agent *does* and things that simply happen to the agent. It seems that if an agent is responsible for anything at all, he is responsible for that which is within his control. This is centrally due to the fact that we have reasons to want our actions to be an accurate representation of our moral commitments. Given that our participation in interpersonal relationships inevitably exposes us to reactive attitudes, we have reasons to want these responses to be reflective of ourselves, and not of factors which interfere with our agency. The concept of control can be invoked to mark this distinction between on the one hand actions which result from my free agency and on the other actions which result from influences alien to my own agency.

But the stakes here are higher. For one’s sense of self, one’s understanding of one’s place within the natural world, one’s understanding of precisely what kind of thing one is within such a naturalistic framework, all of these depend crucially on one’s demarcating the aforementioned boundaries, namely between what happens to one and what one does. In the same way, Randolph Clarke argues that: “The dignity that one has in virtue of being a free agent consists partly in the fact that, in acting freely, one makes a

---

difference, by exercises of active control, to how one’s life goes and to those things that can be and are affected by one’s free actions.”9 According to Clarke, we care about control because of the dignity with which it endows human life. The mark of this kind of agency, the agency in which one has control over one’s destiny, is that it enables the agent to effect changes in the world, in Clarke’s phrase, to make a difference to history.10

It should be clear that the concept of control is capable of serving these two functions only in virtue of its being a causal notion. Possession of control endows agents with certain powers through the exercise of which we become active with respect to our lives. The concept of control is essentially that of the ability to bring about or produce certain effects in the world.11 This can be clearly seen in Dennett’s characterization of control as outlined above.12 Thus if the model airplane which I am commandeering veers out of range, I can be said to have lost control of it precisely because I am unable to bring about or produce certain effects with respect to it. The same conception of the nature of control is also all but explicit in much of the incompatibilist literature.13 On these views, the reason the thesis of determinism is seen to erode control is that both determinism and control are taken to be causal concepts. Thus the truth of determinism is seen to block the ability of the agent to effect certain changes in the world, paradigmatically to act otherwise.14

10 Ibid.
14 This, of course, should not be taken as an endorsement at this stage on my part of any particular view about the sort of control required for moral responsibility. Nor should it be taken as an endorsement of the view that the sort of control required for moral responsibility is causal control. What I am suggesting here
In this section I have argued that the notion of control is primarily a causal concept, and in particular the relation that holds between two entities, one of which is the controller, capable of exercising power over another entity, which is that which is being controlled. I will develop and defend this claim in the course of the following two sections. But in addition to the task of developing and defending this conception of control, I am interested in the implications of this account of control for the voluntarist interpretation of the quality of will thesis. Given this account of intention, is Wallace correct that intention is the chief attitude with which moral criticism is concerned?

4.4 Quality of Will, Intention, and the Morally Significant Attitudes

In order to answer this question, we can begin by recalling Dennett’s example of the father who forgets to drop off his infant at day-care, leaving her strapped in her car seat, which results in her death. I noted earlier that it is a compelling intuition that the range of Strawsonian reactive attitudes constitutive of moral criticism is not appropriate in the case of the father. But the argument of Section 4.3 supplies the resources for challenging this intuition. What is important to the question of whether moral criticism of the father is appropriate is not whether the father in fact intended to kill his infant, but rather whether we can articulate a compelling account of why the father ought not have left his infant abandoned in the car. Of course the issue of intent is immediately relevant when it comes to specifying the degree of resentment and indignation that is warranted. There seems to be a significant difference between a father who forgets about his infant strapped in her car seat, and a father who intentionally abandons his infant. This difference, however,
does not pertain to whether these agents can be appropriately blamed, but rather to the
more specific question of the kind and degree of response that is rendered appropriate by
these agents’ failures.

In order to bring this point into sharper focus, compare Dennett’s example of the
forgetful father with the example of a parent who leaves his or her infant in a car while
running an errand or paying a visit to the local casino. What is crucial to note is that the
proposition that intention is the primary attitude with which moral criticism is concerned
would lead to a counterintuitive conclusion. It commits us to the conclusion that such a
parent cannot be appropriately morally criticized for harming his infant, because it is not
the case that a parent who abandons his infant in a car in order to simplify running
errands intends to harm the infant. Therefore, if intention was the primary (i.e. a
necessary) attitude for moral criticism, such a parent would be an inappropriate candidate
for moral criticism for harming the infant. This conclusion, however, is highly
implausible. The question of what this agent’s intentions were does not seem to make any
immediate difference to our assessment of whether he can be appropriately morally
criticized for harming the infant. What this suggests is that the agent’s intention is not the
only attitude with which moral criticism is concerned.

I want to suggest that the reason for this is the tension between on the one hand
the quality of will thesis’ concern for substantive issues about the content of the agent’s
attitudes, and on the other the formal concern of the voluntarist interpretation in terms of
the existence of the appropriate causal relationships between the controller and the
controllee. In other words, what is distinctive about Wallace’s position is that his
voluntarism arises out of the quality of will thesis. Thus Wallace is committed to at least two claims: (1) that possession of control (i.e. intention, choice) is a necessary condition of appropriateness of moral criticism, and (2) that a crucial feature of moral criticism is concern about the quality of the agents’ attitudes. But on my diagnosis there is a tension between these two claims, as commitment to (1) is precluded by a commitment to (2), and vice versa. Thus while the father who ‘merely forgets’ about his infant secured in the back seat might well be open to moral criticism based on the quality of will thesis, because taking appropriate precautions is “open to the influence of reasons” and therefore a failure to do so is indicative of a failure to pay sufficient attention to the relevant features of the world, the agent in question clearly does not intend to bring the tragedy about. Hence if it is the quality of will thesis which underwrites moral criticism (as the objection concedes) and if it is inconsistent with the central claim of voluntarism (as I have argued), then it follows that Wallace’s voluntarist interpretation of the quality of will thesis is unacceptable.

The immediately preceding discussion might give the impression that cases of negligence or omission are the only ones where Wallace’s voluntarism fails. Immediately following the above-quoted passage, Wallace make just such a claim. Having invoked the quality of will thesis to account for cases of negligence and omission, he writes:

But when neither negligence nor recklessness is at issue, it seems clear that excusing conditions...function to render what one did (or an aspect of what one did) unintentional, where this in turn means that one has not in fact violated a moral obligation after all.15

15 Wallace, Responsibility and the Moral Sentiments, p. 139.
But the problem I have highlighted above becomes even more poignant in cases of excuses and exemptions in general. Take for example, the case of someone who is *serially* guilty of becoming inebriated and making racially offensive jokes. Let us assume that the agent in question did not *intend* to undermine the dignity of those individuals who are singled out by the offensive joke. That is to say, he did not *mean* to insult or degrade anyone. Rather, in his inebriated state, he failed to take into account the relevant considerations and therefore failed to see that the joke is offensive. But such a person is clearly not entitled to an excuse. Were Wallace correct in the claim that excuses track whether the agent acted intentionally, the serial racist would be entitled to an excuse in each episode of telling a racist joke under the influence of alcohol. But this clearly seems false. We judge him responsible because he continues to display the offensive actions, thereby betraying a morally faulty quality of will. Indeed, it is perhaps the very fact that he continues to display the same behavior when drunk, and when certain socially instituted and maintained cognitive filters of his are not operative (for example, when, due to being inebriated, he has lower regard for standards of etiquette, or diminished concern about what others think of him, etc.) that give us compelling reasons to take his conduct to be a sign of a faulty quality of will. This reinforces the conclusion reached earlier, namely that Wallace unjustifiably restricts the sort of attitudes implicated in the quality of will thesis to intentions.

I have argued that Wallace’s voluntarism fails to offer a plausible interpretation of the quality of will thesis. This failure is initially most poignant in cases of omission and negligence. However, the tension I isolated between Wallace’s over-arching voluntarism
on the one hand and his appeal to the quality of will thesis on the other leads me to conclude that the failure of his voluntarism also permeates Wallace’s entire account of excuses. This suggests that the importance attached to the requirement of intention (and the condition of control generally) is misplaced. When we excuse agents from moral criticism, this is*not* done, in principle, because they lacked control over their actions. By the same token, when we *do* hold them accountable for their conduct, this is not done because they did have control over their actions.

We may pause at this point to consider an objection to the jettisoning of the notion of control as a necessary condition of moral criticism. If control demarcates the boundary between things an agent does and things that happen to an agent, then rejecting the notion of control is in effect to obliterate the notion of responsible agency altogether. If we abandon the concept of control, then we would be unable to distinguish, on principled grounds, between things an agent does (actions that are under his control) and things that happen to him (actions that are not under his control). Given that if we are responsible at all, we are responsible for things we do, to reject control is either to shrink the sphere of responsibility infinitesimally, or expand it infinitely. The consequence of rejecting control would be either that we are just as responsible for things that happen to us as things that we do, and therefore responsible for everything, or else some sort of skepticism according to which rejection of control creates a normative vacuum making moral criticism entirely and altogether inappropriate.

Thus it may be argued that the conclusion that we must abandon the condition of control is too extreme. We need, instead, to *refine* the condition of control by specifying
the sorts of control that we deem essential to free agency. Wallace, for example, has argued that in cases when we exempt agents from ascriptions of responsibility, we do so because the agent lacks what he calls reflective self-control.\textsuperscript{16} According to Wallace, reflective self-control consists, in the first place, of the power to grasp and apply moral reasons, and, in the second place, of the power to control and regulate behavior by the light of such reasons.\textsuperscript{17} On Wallace’s account, therefore, reflective self-control requires both cognitive and affective powers.\textsuperscript{18} The power of grasping and applying moral reasons is essentially a cognitive power, involved in understanding moral principles and extending them to a wide variety of cases. The power to regulate and control one’s behavior in light of such principles, on the other hand, concerns the ability to translate one’s convictions with regards to the moral weight of these principles into action.\textsuperscript{19}

A similar view is suggested by McKenna, who, having conceded that voluntarism, in its original version, is not sustainable, goes on to outline what can be considered a normative interpretation of the condition of control. According to this view:

An agent can be morally responsible for her attitudes, character traits, or any other nonvoluntary object of responsibility only if [it] stands within the scope of her rational control. Here, rational control has two components. One involves the possibility of rational activity.... A second involves a standing capacity to perform a free mental act of deciding or choosing to evaluate one’s moral standpoint(s).\textsuperscript{20}

According to McKenna, this characterization of the nature of free agency is clearly a concession on the part of the voluntarist, for it does reject the strong voluntarist claim that

\begin{itemize}
\item \textsuperscript{16} Wallace, \textit{Responsibility and the Moral Sentiments}, p. 166.
\item \textsuperscript{17} Wallace, \textit{op. cit.}, p. 157.
\item \textsuperscript{18} Wallace, \textit{op. cit.}, p. 159.
\item \textsuperscript{19} Wallace, \textit{op. cit.}, p. 158.
\end{itemize}
the condition of control must be fulfilled in every episode of morally responsible action. Therefore, one may be responsible for an attitude, or a character trait, even if such an attitude or character trait is not under the control of the agent, nor can it be traced to an earlier choice of the agent. But, according to McKenna, this is only a partial victory for the non-voluntarist, for his view still requires that control figure somewhere in our assessments of the agent. In cases of responsibility for the non-voluntary, control comes in “indirectly, via a (sometime unexercised) capacity to decide freely to evaluate one’s moral standpoint (as regards the relevant nonvoluntary object of responsibility)”\textsuperscript{21}

In this way, both McKenna and Wallace seek to retain the requirement of control, but offer a more plausible conception of the kind of control required for appropriateness of moral criticism. However, this strategy is vulnerable to two related objections. First, they have in effect merely insisted on the claim that control is a necessary condition of appropriateness of moral criticism. But the issue is \textit{not} whether some conception of control or another is capable of solving the philosophical puzzle about the relationship between control and the appropriateness of moral criticism. Rather, the relevant question is, having encountered the failure of the causal conception of control, why must we continue to accept the proposition that control, in some relevant sense, is a requirement of moral criticism.

Furthermore, it is not clear in what \textit{further, non-causal} sense the capacities implicated in Wallace’s and McKenna’s theories are entitled to be considered species of control at all. The normative conception of control offered by Wallace and McKenna does not capture a freedom-relevant condition, but rather designates the kind of normative

\textsuperscript{21} McKenna, “Putting the Lie on the Control Condition for Moral Responsibility,” p. 36.
relations that must hold between the agent and the reasons which apply to him. In this sense, it is not clear precisely why we should accept the presentation of the sort of normative competence highlighted by Wallace and McKenna as control. Recall that on McKenna’s account, the rational control required for appropriate moral criticism consists in part in the possession of “a standing capacity to perform a free mental act of deciding or choosing to evaluate one’s moral standpoint(s).” It is difficult to understand what his capacity involves other than the power and ability to bring certain effects into the world, that is, of exercising control over one’s desires and commitments. McKenna seems to be incorporating an agent-causal conception of the sort of capacity required for appropriateness of moral criticism. He therefore is unable to distance himself from the failures of the voluntarist interpretation of the quality of will thesis identified earlier.

A similar criticism applies to Wallace. Recall that, on Wallace’s account, one’s obligations towards others is not only that one should not make bodily movements likely to inflict harm on others, but rather that one should not make choices or decisions that express an indifference to others. As I have noted in the previous chapter, this is because on Wallace’s account:

[Conditions of accountability] must make it fair to demand that [some agent] s comply with moral obligations we accept, where the obligations are supported by distinctively moral reasons. But it would seem fair to demand this of a person only if the person possesses what I shall refer to as the powers of reflective self-control: (1) the power to grasp and apply moral reasons, and (2) the power to control or regulate his behavior by

---

22 Robert Kane argues along the same lines. He holds that evaluations of Wallace’s notion of reflective self-control must focus on precisely what sort of control is being implicated. More specifically, he argues, as I have done, that to have control is, in essence, to have the power to determine whether or not a certain fact occurs. But this is clearly not what reflective self-control designates. See his “Responsibility, Reactive Attitudes and Free Will: Reflections on Wallace’s Theory,” in Philosophy and phenomenological Research 64 (2002) pp. 693-8, at p. 698.
Therefore being an appropriate candidate for moral criticism requires not only the
cognitive capacity to grasp moral norms, but also the executive power of controlling and
regulating one’s conduct in light of one’s understanding. Clearly, this itself requires some
form of control by the agent, that is, the causal power to bring certain effects into the
world.

In this section I have argued that intention is not the primary attitude with which
moral criticism is concerned, and defended this claim against the competing normative
interpretation of control advanced by Wallace and McKenna. More specifically, I argued
that since control is a causal concept, the voluntarist interpretation of the quality of will
thesis is incapable of addressing substantive concerns about the content of the agent’s
attitudes. This creates a conceptual gulf between the quality of will thesis and condition
of control (and, in our case, the requirement of intention). The question I will investigate
in Section 4.5 is whether the notion of tracing can bridge this gulf. I argue that it cannot.

4.5 Control, Tracing, and Responsibility

In Section 4.4 I argued that intention is not the primary attitude with which moral
criticism is concerned. It may be objected that this conclusion relies on an inadequate
analysis of the complexities of the agent’s motivational history. According to this
objection, while it is correct that in a father who intentionally leaves his infant in the car
to simplify running errands does not intend to harm the infant, he does intend to leave his
infant in the car. It is, furthermore, a reasonably foreseeable consequence of this act that
the infant may be harmed. Therefore, it may be objected, I have failed to show that

intention is not the primary attitude with which moral responsibility is concerned.

This objection raises important issues about the connection between tracing and appropriateness of moral criticism. In this section I will advance two arguments. First, I argue that Wallace’s argument, and the objection under consideration in this section, rely on the assumption that the tracing of some unintentional harm to an earlier choice is always possible. This claim however, is false. Second, I argue that even if it is in principle possible to explain all cases of negligence and omission by reference to an earlier intention, this would not allay the difficulties with the voluntarist interpretation of the quality of will thesis.

Let us begin, however, with a brief discussion of the requirement of tracing, and its significance for the theory of responsible agency. According to Manuel Vargas,

Tracing is the idea that responsibility for some outcome need not be anchored in the agent or agent’s action at the moment immediately prior to outcome, but rather at some suitable time prior to the moment of deliberation or action.\(^{24}\)

The thought behind the notion of tracing is that an agent can be appropriately morally criticized even if he does not satisfy some specified condition of moral criticisms at the time of the infraction, provided that he does satisfy the requirement in question at some earlier time, where there exists an appropriate relationship between the agent when he does not satisfy the relevant requirements, and the earlier time when he did. In order to make this clear, we can reconsider Wallace’s attempt to allay difficulties cases of

---

negligence and omission are seen pose for his voluntarist account of conditions of responsible agency:

Negligence and forgetfulness are slightly harder cases, [than cases of recklessness] perhaps, because [contrary to the cases of recklessness] there may not even be awareness of the risks involved at the time when one acts negligently or forgetfully. Here one may have to trace the moral fault to an earlier episode of choice: agreeing to look after the child, for instance, I may then have failed to take steps sufficient to ensure that I would uphold this agreement, where this led to the child’s negligently being harmed. Once again, this feature of my earlier choice – whether or not I took precautions to ensure that my agreement would be carried out – is subject to the direct influence of reasons since I might have chosen to take precautions because I recognized this to be required by my agreement to care for the child. In this way, negligence and forgetfulness may also be traced to a blameworthy quality of will.\footnote{Wallace,} 25

According to Wallace, even granting that negligent and forgetful harms are characterized by an absence of intention, agents can nevertheless be appropriately morally criticized for them provided the negligent or forgetful acts can be traced to an earlier episode of choice.

However, Wallace’s argument suffers from two flaws. In the first place, Wallace seems to suggest that the following two possibilities are exhaustive of all cases in which the question of appropriateness of moral criticism arises: (1) the agent intentionally performs the action, or (2) the unintentional consequences can be traced to an earlier intentional action. But while there certainly \textit{are} cases in which such tracing of the agent’s present state to prior episodes of choice can be done, it seems equally clear that it is not necessarily the case that this can be done in all cases of morally culpable omission and negligence. From the fact that there are cases in which our tendency to morally assess agents for their omission of forgetfulness can be traced to an earlier choice it does \textit{not} follow that this can be done in every such case. Agents may be open to moral criticism.

\footnote{Wallace, Responsibility and the Moral Sentiments, p. 139.}
even if we are unable to provide a plausible narrative which explains their culpable omission or forgetfulness in virtue of a prior choice.

In order to substantiate this claim, consider Thomas Nagel’s influential discussion of moral luck.\textsuperscript{26} Nagel invites us to consider the case of a driver who runs over a child:

The driver, if he is entirely without fault, will feel terrible about his role in the event, but will not have to reproach himself. Therefore this example of agent-regret is not yet a case of moral bad luck. However, if the driver was guilty of even a minor degree of negligence — failing to have his brakes checked recently, for example — then if that negligence contributes to the death of the child, he will not merely feel terrible. He will blame himself for the death. And what makes this an example of moral luck is that he would have to blame himself only slightly for the negligence itself if no situation arose which required him to brake suddenly and violently to avoid hitting a child. Yet the negligence is the same in both cases, and the driver has no control over whether a child will run into his path.\textsuperscript{27}

There are two ways in which we may understand this example. In one case, the agent fails to appropriately inspect and maintain his vehicle out of laziness, pure indifference, or extraordinary and unreasonable concern with saving money. Such a person decides that taking steps to inspect his vehicle involves too much effort or money, and elects to take a calculated risk. In the second case, by contrast, the agent fails to have his vehicle maintained appropriately not out of laziness or calculation, but rather because he ‘merely forgets.’ For example, we can imagine that preparing for an important upcoming project has occupied the agent to such a degree that he has been out of touch with the ordinary, day-to-day concerns, such as maintaining a clean house, keeping up with his bills, and having his car appropriately maintained. Such an agent does not choose not to fulfill these obligations. These concerns simply become of secondary importance given the

\textsuperscript{27} Nagel, op. cit., p. 178, emphasis in the original.
weight he attaches to his pending project.

While both of these situations are cases of moral luck, only the latter shows the shortcoming of Wallace’s appeal to the notion of tracing. This is because in the first case, the fact that the agent is involved in an accident can be traced to an earlier choice, namely the choice (decision) to refuse to maintain his car as required and face the consequences. By contrast, in the second case the failure of the agent to maintain his vehicle properly, a failure which leads to accident, satisfies neither of Wallace’s two conditions. First, it is clear that, as the scenario is articulated, the driver does not intend to run the child over. Furthermore, and perhaps more importantly, his unintentional running over of the child is not traceable to an earlier intention.

However, if this assessment of the example is correct, Wallace’s appeal to tracing fails. This is because while the driver has been merely negligent, he still seems to be an appropriate candidate for moral criticism. Of course, the kind and degree of reactive attitudes that can be appropriately directed at this agent is importantly different from the reactions that would be warranted had he intended to cause the harm in question, or, more to the point, if his unintentional causing of the harm could be traced to an earlier choice, but this difference is not that of appropriateness of moral criticism. Since the harm is neither intentional, nor can it be traced to an earlier intention, and since the agent can be appropriately morally criticized for it, it follows that Wallace’s tracing condition is false.

Thus far in this section I have argued that there are harms for which an agent can be appropriately morally criticized, even if the harm is neither intended nor traceable to
an intention. In the remainder of this section I argue that even if some harm could be traced to an earlier intention, this would still not provide a compelling voluntarist account of conditions of moral criticism.

The primarily difficulty here is that tracing some unintentional harm to an earlier intentional action gets the content of moral criticism wrong. Consider again the variation on Dennett’s case in which a father leaves his infant in the car while running errands. Even if we grant that the harm that comes to the infant can be traced to an earlier intention, by Wallace’s own lights this makes moral criticism of the father appropriate only in regards to his act of abandoning his infant in the car, and not for his unintentional consequences of this action. Even if we accept Wallace’s contention that intention is the primary attitude with which moral criticism is concerned, and even if we can articulate a plausible account of how an unintentional harm can be traced to an earlier intentional decision, this would provide a normatively adequate basis for blaming the agent for the initial intended action, and not necessarily for the unintended consequences. This distinction is particularly important because the range of attitudes warranted by the consequences of the action are radically different from the reactive attitudes warranted by the initial intended infraction.

Wallace therefore owes an account of what allows the culpability for the antecedent choice to transfer to a subsequent unintended harm. Granting that all cases of omission or negligence can be traced to an earlier episode of choice, the question is how can this earlier choice make the agent responsible for the some unintentional states of affairs. We can appreciate the problem such a transfer poses, and the solution it calls for,
by considering what Fischer calls the principle of transfer of nonresponsibility.\(^{28}\)

According to this principle, which constitutes the direct argument for incompatibilism,\(^{29}\) if no one is responsible for a certain fact, \(p\), and no one is responsible for the fact that \(p\) entails some consequence, \(q\), then no one is responsible for \(q\). We are, by contrast, in need of a principle of transfer of responsibility, a principle that explains how an agent who is responsible for a choice can be responsible for the unintended consequences of that choice.

One way this can be done is by appealing to the epistemic constraints on moral criticism. We can say that if a certain consequence is a *foreseeable* result of a certain prior intentional action or choice, then one’s responsibility for the action in question can transfer to a later consequence. The articulation of a systematic account of this epistemic condition is beyond the scope of the present discussion. It will suffice for our present purposes to say that the condition of foreseeability can be either subjective or objective. That is, it can either require that a certain consequence be foreseeable from a subjective point of view, where we adopt the point of view of the agent by entering the agent’s epistemic standpoint, or it could require that a certain consequence be foreseeable from an objective, perspective-independent standpoint.\(^{30}\)

Is the epistemic constraint capable of bridging the gap between the quality of will thesis and the requirement of control, such that an agent can be seen to have acted intentionally in virtue of having known the likelihood of a certain action bringing about


\(^{30}\) Thomas Nagel, *The View from Nowhere* (Oxford University Press, 1986) p. 120.
bad consequences (even if the agent did not intend to bring about those consequences)?

George Sher has recently advocated an epistemic conception of control. According to Sher, an agent can lack control over his actions in virtue of the failure of his non-executive faculties. Among such non-executive faculties whose failure can erode control Sher includes epistemic faculties. Under such circumstances, while the agent does exercise causal control and brings about certain effects in the world, the fact that he fails to note some salient fact about his situation suggests that he lacks appropriate control over it. Consider the following example:

*Home for the Holidays.*—Joliet, who is afraid of burglars, is alone in the house. Panicked by sounds of movement in her kitchen, she grabs her husband’s gun, tiptoes down the stairs, and shoots the intruder. It is her son, who has come home early for the holidays.

In such a case, while Joliet does possess the sort of ability constitutive of the causal interpretation of control, there is a sense in which she is acting based on incomplete information and therefore lacks control over her the incident.

Sher is interested in enumerating ways in which epistemic failures undermine the condition of control. For now, I am not interested in the question of whether the requirement of control is defensible. Rather, supposing that it is, I want to suggest that Sher’s argument has the impact of strengthening the condition of control by supplementing the merely causal understanding of control with an epistemic understanding of control. On this understanding, for our exercises of power to be capable

32 Sher, op. cit., p. 288.
of grounding moral responsibility, they must be based on an accurate picture of the world. The failure of agents to satisfy this epistemic condition means, according to Sher, that the agent’s exercises of power are, in effect, futile.\textsuperscript{34} On Sher’s account, for an action to be voluntary in the appropriate sense, the agent must not only be capable of exercising causal control, but also must satisfy certain epistemic conditions.

Sher’s analysis therefore gives rise to three related questions. The first question is whether moral criticism is appropriate only if the agent satisfies some specified epistemic requirement. In this connection Sher argues that the fact that Joliet seems an appropriate candidate for moral criticism suggests that the traditional view according to which the satisfaction of some appropriately specified epistemic requirement is mistaken.\textsuperscript{35} The second question concerns the broader issue of the connection between voluntariness of action and appropriateness of moral criticism. In this connection, Sher argues that cases such as Joliet illustrate that agents may be appropriately morally criticized even if they fail to satisfy the requirement of control (because they can be appropriately morally criticized even if they do not satisfy the epistemic constraint, which he takes to be a requirement of control). While I think Sher’s answers to both of these questions are correct, I also think that there is also a third question that deserves attention. This is the question of the status of the epistemic constraint on moral criticism, and in particular whether Sher is correct to construe the epistemic constraint as concerned with the agent’s

\textsuperscript{34} See Laura Waddell Ekstrom, \textit{Free Will: A Philosophical Study} (Westview Press, 2000), p. 211 for a similar view.

\textsuperscript{35} See George Sher, \textit{Who Knew? Responsibility without Awareness} (Oxford University Press, 2009), where he combats what he calls the “searchlight view” of control and moral criticism, according to which the awareness of the relevant aspect of one's situation is a necessary condition of the satisfaction of the requirement of control and therefore satisfaction of conditions of appropriateness of moral criticism (pp. 5-6).
control. This third question is significant for at least two reasons. First, unless the epistemic constraint can be appropriately construed as a requirement of control, the crucial link between the first and the second questions would be missing. In other words, unless the epistemic constraint is a requirement of control, it would not follow from the fact that Joliet can be appropriately morally criticized in spite of failing to satisfy the epistemic constraint that the broader claim concerning the necessity of control for the appropriateness of moral criticism is false. Is Joliet really a counterexample to the proposition that moral criticism requires control?

The question of the relation between the requirement of control and the epistemic constraint is also significant for the related reason that defenders of the requirement of control often appeal to the non-causal interpretation of it to account for certain problematic cases. On these accounts, even if one grants that the purely causal conceptions of control are rendered implausible by certain counterexamples, it nevertheless remains true that control in some appropriate, non-causal sense is a necessary requirement of appropriateness of moral criticism. This invites the question of whether a non-causal account of control, along the lines suggested by Sher, is acceptable.

There are two reasons for answering this question in the negative. First, it is important to note that the condition of control is a “freedom-relevant” constraint on moral criticism.\(^{36}\) It singles out the sort of circumstances which interfere with the agent’s free exercise of power over his environment. Sher’s epistemic condition, by contrast, does not seem to be immediately concerned with power at all. It is concerned, rather, with the degree to which one exercises one’s power based on correct information about the

circumstances. It therefore is not difficult to see that one may well satisfy the freedom-
relevant condition, that is, one’s actions might be “free” in the sense we associate with
moral responsibility, but fail to satisfy the epistemic requirement. One may act freely
(and therefore responsibly) without having acted correctly. Joliet acted wrongly, in the
sense that her actions were not warranted by features of the world. But not every wrong
action is unfree in the relevant sense. After the incident it would be odd for Joliet to claim
that she somehow lacked control over her shooting her son, that her action was somehow
unfree. The apt response would be to say “I over-reacted.”

I have argued that the epistemic constraint is incapable of capturing the
requirement of control’s concern with the notion of free agency. A similar conclusion can
be reached in connection with the concern with the notion of attributability. Recall that as
I have noted earlier, the question of applicability of excuses is centrally a question about
the conditions of attributability of actions. For an agent to claim that some event was not
due to his choice and intention is for him to claim that it was not his action (if an action at
all), that the action was not attributable to him to begin with. In such cases, the plea of “it
wasn’t me” is forceful and, even if it is ultimately to be dismissed, deserve consideration.
But it is not clear whether the failure of an agent to satisfy the epistemic constraint
entitles him to such a plea. In the case of Joliet, for example, while the fact that she acted
under ignorance and duress mitigates our reactions towards her, it is not clear that the
action of her shooting her son is in fact not attributable to her.

I have argued so far that the epistemic constraint is not a control-related
constraint. This claim exposes a a tension within Wallace’s voluntarist interpretation of
the quality of will thesis. This tension is between the epistemic articulation of the principle of transfer of responsibility and the requirement of control. The difficulty for Wallace is that relying on this epistemic condition weakens the freedom-relevant aspect of his voluntarism. The fact that some state of affairs is a foreseeable consequence of some intentional action cannot provide a basis for morally criticizing the agent for the unintentional consequence, unless we reject the claim, essential to the voluntarist interpretation of the quality of will thesis, that intention to perform some action is a necessary condition of being blameworthy for that action.

The employment of the epistemic condition is therefore damaging to Wallace’s voluntarist interpretation of the quality of will thesis precisely because it undermines his commitment to the requirement of control. Wallace cannot consistently maintain that choice and intention are necessary conditions of appropriateness of moral criticism, but then specify an epistemic conception of the transfer of responsibility that, as I have argued, is not a causal and therefore control-related requirement at all.

4.6 Concluding Remarks

In this chapter I have investigated the question of whether there is a mens rea constraint on moral criticism, that is, whether an agent is an appropriate candidate for moral criticism only if he intends to bring about some harm. I focused on Wallace’s voluntarist interpretation of the quality of will thesis. According to this interpretation, intention is the chief attitude which attracts moral criticism, and an agent is an appropriate candidate for moral criticism only if the agent’s action expresses a choice or intention to fail to uphold the requirement of good will and due regard. I argued that this
view encounters serious difficulties accounting for cases of negligence and omission.

In Section 4.3 I argued that intention is essentially and primarily a causal notion. Therefore, as I argued in Section 4.4, it is incapable of providing a robust basis for all of our practices of moral criticism. More specifically, I argued in Section 4.4 that indifference and negligence typically bring about harmful consequences precisely in the absence of an element of intention, but we continue to think of such agents as open to moral criticism.

In Section 4.5 I turned my attention to the question of the connection between tracing and the quality of will thesis. On this view, even if an action of the agent does not issue from an immediate choice or intention of the agent, the agent may appropriately be held accountable in so far as the action in question can be traced to an earlier choice. In response to this argument, I defended two claims. I first argued that agents can be appropriately morally criticized for harms that are neither intended nor capable of being traced to an earlier intention. I then argued that even if some unintentional harm is traceable to an earlier intention, this would not help Wallace’s voluntarist interpretation of the quality of will thesis. This is because this would provide an inaccurate account of the content of moral criticism.

I conclude that there are important conceptual difficulties in providing a voluntarist interpretation of the quality of will thesis. In Chapter Five I will turn my attention to T. M. Scanlon’s contractualist interpretation of the quality of will thesis. This is important for two reasons. In the first place, the contractualist interpretation offers the resources to diagnose and remedy the attractiveness of the requirement of control.
Furthermore, it provides a compelling account of the connection between moral criticism and reasons. My main aim in Chapter Five will be to argue that contractualist resources allow for a plausible interpretation of Strawson’s project in “Freedom and Resentment,” and of the quality of will thesis in particular.
CHAPTER 5: FREEDOM, FAIRNESS, AND REASONS

5.1 Introduction

In the previous chapter I argued that the voluntarist interpretation of the quality of will thesis is false. Many philosophers have likewise advanced arguments for the conclusion that the condition of control is false.¹ In spite of these arguments, the condition of control retains much of its intuitive and philosophical appeal. Many philosophers, whether compatibilist or incompatibilist, hold that the philosophically interesting question is not whether warranted moral criticism demands the satisfaction of the condition of control, but rather what kind of control and how much of it.

The persistence of the condition of control therefore requires both diagnosis and remedy. There are, more specifically, two questions that are of importance. The first question is why the condition of control continues to enjoy the appeal it does. The second question is whether we can articulate a normatively plausible account of preconditions of moral criticism that does not appeal to the condition of control.

One reason for the appeal of the condition of control is that control is taken to specify the freedom-relevant constraint on moral criticism. Voluntarist accounts rely on what we can call the up-to-us-ness intuition. Even if one finds particular voluntarist constraints on moral criticism unacceptable, it seems difficult to deny the proposition that being blameworthy requires freedom of one sort or another, that is, requires that some

crucial aspects of the agent’s psychology and circumstances be up to him. According to
Randolph Clarke, for instance,

The dignity that one has in virtue of being a free agent...consists in one’s
making a difference, by one’s exercises of active control, to how things go
in the world. It consists in one’s actions’...being attributable to one as their
source and author, and, provided that one has an ordinary capacity to
appreciate and act for moral reasons, in one’s being responsible for one’s
actions.... And it consists in one’s not being routinely subject to a
practically unavoidable illusion regarding the openness of alternatives to
one’s acting as one, in fact, does.²

There are at least two ideas in this passage that deserve emphasis. The first is that the
ideals of control and freedom (i.e. autonomy) are necessarily related. To be in control of
one’s actions just is to satisfy some freedom-relevant condition. The second idea is that
the satisfaction of some specified condition of autonomy and freedom is essential to the
attainment of the integrity constitutive of (or at least necessary for) personhood. Part of
this dignity is that persons are taken to be appropriate objects for a range of attitudes
constitutive of moral responsibility, attitudes that are unwarranted towards non-persons.
On this account, therefore, moral criticism is appropriate only if the condition of
autonomy is satisfied. What this shows is that the ideals of freedom and autonomy are
essential in a plausible account of conditions of moral criticism. The requirement that one
must have control over one’s actions or attitudes in order to be appropriately subject to
moral criticism seems a natural way to specify the freedom-relevant condition.

Is this account of conditions of responsible agency correct? Is it true that the
ideals of freedom and autonomy are essential to a normatively adequate conception of
preconditions of moral criticism? In this section I argue that it is not. My argument will

proceed as follows. I will begin, in Section 5.2, by articulating a contractualist account of conditions of responsible agency, owing to T.M. Scanlon. On Scanlon’s account, obligations we have to other persons arise from substantive reasons that legislate in favor of (or against) treating agents in certain ways. These reasons arise from the value of persons as rational beings, and bear directly on the quality of our interpersonal relationships. The crucial question for the proponent of the condition of control (and the attendant requirement of autonomy), therefore, is what further condition is missing from Scanlon’s contractualist account?.

Recently some philosophers have argued that Scanlon fails to dissociate moral criticism from freedom. R. Jay Wallace, for example, focuses on the case of wayward desires, desires that resist the agent’s attempts to modify or withdraw them, and argues that to morally criticize the agent for such desires is to place an unfair demand on the agent. In Section 5.3 I argue that this claim can be understood in either of two ways. It may be understood as a claim about the accuracy of moral criticism. On this account, to morally criticize an agent for a wayward desire is to misidentify the target of moral criticism. Alternatively, it could be understood as a claim about the appropriateness of imposing on the agent the set of burdens (such as censure and isolation) that the agent cannot, as a matter of fact, escape.

In this chapter I argue that both of these claims are mistaken. In Section 5.4 I consider the latter claim. I argue that once we distinguish between the question of preconditions of moral criticism and the question of the content of moral criticism that is rendered appropriate by the agent’s fault, the worry about the cost of moral criticism for
recalcitrant desires loses its force. In Section 5.5 I consider the claim about the inaccuracy of moral criticism for wayward attitudes, and argue that since moral criticism primarily concerns the justifiability of an attitude in light of reasons, it would not be inaccurate to demand justification from an agent for wayward attitudes. The justifiability of an attitude, and therefore appropriateness of moral criticism, are independent from the ability of the agent to control those attitudes.

In concert, these two arguments provide compelling reasons for rejecting the requirement of freedom, without also jettisoning the responsible agency essential to human integrity.

5.2 Judgment-sensitivity, Justifiability and Impairment

Scanlon distinguishes between two ways in which judgments of responsibility, that is, propositions of the form “x is responsible for y,” where x is an agent and y an action or attitude, are employed: substantive responsibility and responsibility as attributability. Substantive responsibility concerns what a person’s obligations are, i.e. the ways in which he is required or not permitted to act. This question can be asked within a narrow and a broad scope. Consider the example of smoking. Within the broad scope of substantive responsibility, we can ask whether a person is responsible for not smoking in a public place. In asking such a question, we are inquiring as to whether it is required of a person not to smoke, or, alternatively, whether is permissible for him to smoke in a public place. Suppose that someone chooses to continue smoking in spite of knowing about the

---

4 Ibid.
adverse health consequences that accompany the habit. In this case we can pose the substantive question in a narrow sense, viz. whether it is the agent himself that ought to bear the burdens of the consequences of his smoking, namely his poor health and the cost of treatment, or whether he can reasonably expect others to relieve him, either fully or partially, of the said burdens.

Judgments of moral responsibility can also be employed in a different sense, which Scanlon calls responsibility as attributability. An agent is responsible in this sense only if it is appropriate to take his action or attitude as a basis for his moral appraisal. Judgments of responsibility as attributability are silent on the question of the nature (i.e. the content) of the moral appraisal that the action in question warrants, and focus instead on the question of whether there exists an appropriate connection between the agent and his actions, such that the action can be taken as a basis for appraising the agent.

In order to illustrate the distinction between these two employments of judgments of responsibility, consider the example of a bank-teller who is forced to choose between the one hand saving his own life and those of all the customers held hostage in the bank, and on the other handing robbers the key to the safe.⁶ The question that is immediately striking, both from the point of view of the teller who is faced with a decision as well as those merely reflecting on the scenario, is what the teller should do. This question calls for a substantive judgment about what this agent’s moral obligations are. In this specific case, as in all cases of deliberation, the agent has to choose between conflicting courses of action and decide between disparate loyalties. Thus the teller might find it unpalatable, to say the least, to accede to the demands of the robbers, not only

---

⁶ The example is from Scanlon, What We Owe to Each Other, p. 279.
because in doing so he would be setting a dangerous precedent (in the sense that it suggests that robbers can gain access to the bank’s money by threatening the lives of the bank’s employees and customers) but also because in acceding to the robbers’ demands, the teller might feel that he is letting his employer down by failing to safeguard the bank’s assets. At the same time, given that the teller has reason to take the robbers’ threats seriously, it may seem that endangering many lives is simply not worth the potential loss of assets to the bank. The question of substantive responsibility is whether it is permissible for the teller to allow robbers access to the safe in exchange for the safety of everyone in the bank. The answer to this question requires an account of the teller’s duties and obligations under the circumstances.

By contrast, the question of the teller’s responsibility in the sense of responsibility as attributability does not require an account of the agent’s substantive obligations. Instead, the question of responsibility as attributability is whether the agent’s actions are an appropriate basis for moral appraisal of him, whether the agent is an appropriate candidate for moral appraisal on the basis of what he has done, whatever that appraisal might be.

Scanlon’s thought is that these two kinds of responsibility (i.e. substantive responsibility and responsibility as attributability) have different moral roots. Responsibility as attributability concerns the question of whether an agent is an appropriate candidate for moral appraisal. Therefore, answering this question requires an independently plausible account of the preconditions of moral appraisal. Questions of substantive responsibility, by contrast, require an account of principles specifying duties

---

7 Scanlon, *What We Owe to Each Other*, p. 249.
that we owe to each other, and costs that agents can legitimately be expected to bear as a result of their actions. What is crucial to note here is that the question of whether an agent is responsible in the former sense (i.e. in the sense of responsibility as attributability) can, at least in principle, be settled independently of questions about his substantive responsibilities. In other words, whether an action is attributable to an agent in ways that would make him an appropriate candidate for ascriptions of moral responsibility is independent of questions about what he should have done in the relevant circumstances.

I am primarily interested in Scanlon’s account of responsibility as precondition of moral appraisal. More specifically, I am interested in the question of what are, on Scanlon’s account, necessary and sufficient conditions for the appropriateness of moral appraisal of the agent. As I have argued earlier, Strawson notes:

...the very great importance that we attach to the attitudes and intentions towards us of other human beings, and the great extent to which our personal feelings and reactions depend upon, or involve, our beliefs about these attitudes and intentions. 8

He goes on to note the extent to which:

...we actually mind, how much it matters to us, whether the actions of other people—and particularly of some other people—reflect attitudes towards us of goodwill, affection, or esteem on the one hand or contempt, indifference, or malevolence on the other. 9

According to Scanlon, the fact that we place such an importance on the attitudes others

9 Strawson, op. cit., p. 63.
have towards us, that is, the fact that we are not indifferent about the attitudes that others display towards us, can be highlighted by the implications these attitudes have for our relationship with others.\textsuperscript{10}

Consider the example of someone who hates the game of golf, on the grounds that it is conducted at a slow pace, and is a sport that is typically played by a particular segment of the society, those with the means to afford not only the equipment, but also the time required to practice it. The fact that the agent in question hates this game suggests that he fails to recognize certain features of the game, and the values that these features capture. For example, while the game in question is played at a slow pace, it nevertheless possesses an elegance few other sports do. And while it is typically practiced by a relatively small segment of the society, the range of skills required to succeed in it, the unpredictability of the game itself, and beauty of the landscape in which it is played, are capable of providing enjoyment to all discerning viewers of the game. The fact of this agent’s failure to recognize these values restricts the range of interactions he can have with others, and in particular with those who do have an appreciation for the value of the game of golf. This is principally because he is incapable of sharing in the value of the activity in question. If someone who is an avid fan of golf is confronted with someone who denies the value of the game altogether, the two are bound to feel alienated from each other.

Someone who hates the game of golf is open to appraisal for failing to appreciate certain valuable features of the game. Nevertheless, it is important to note that this does not amount to moral criticism of the agent. It would be inappropriate to blame someone

\textsuperscript{10} Scanlon, \textit{What We Owe to Each Other}, pp. 158-9.
for his dislike of the game of golf, to resent or admonish him, or to be indignant towards him. We can therefore say that such a person does not satisfy a necessary condition of the appropriateness of moral criticism. By contrast, the amoralist, someone who is wholly unmoved by moral norms and the legitimate claims others have against him, attracts the kinds of reactions constitutive of moral criticism. The reactions we have towards the amoralist (reactions such as resentment and indignation) have implications for our relationship with him beyond mere failure to benefit from sharing an activity. The sense of alienation we feel towards the amoralist is not only qualitatively different, but it also results in a more extensive impairment of our relationship with him.

A plausible account of conditions of moral criticism ought to provide an explanation of the grounds for such a difference. The question here is what feature of moral criticism makes it the case that it has such remarkable consequences for the agent’s relationship with others? As Scanlon notes, human beings have the distinctive capacity to assess reasons. According to Scanlon, each human being is “a locus of reasons,” essentially rational and capable of not only assessing reasons and engaging in a process of justification, but also capable of acting on the basis of these reasons. Reason, in the sense employed by Scanlon, is a two-place normative primitive, specifying that certain considerations count in favor of certain attitudes. In the case of the game of golf, for example, the fact that only individuals from a certain socio-economic class are capable of participating in it may be a reason that justifies one’s dislike of the game. By contrast, the finesse the game in question requires and the elegance it exhibits, are facts that count in favor of forming a positive judgment about it (that is, are reasons for liking it).}

---

11 Scanlon, *What We Owe to Each Other*, p. 105.
moral norms are similarly constituted. For example, the fact that driving at excessive speeds and gratuitously failing to observe traffic regulations put the lives of others at risk are good reasons for not doing so.

The fact that persons are loci of reasons makes it the case that the values that are at stake in moral criticism are importantly different from other values. The fact that human beings are rational agents, capable of assessing reasons, of receiving, understanding and offering justification, is itself a reason that counts in favor of treating them in certain ways. More specifically, it seems that humans are entitled to such respect and deference that other living and even conscious beings are not. On Scanlon’s account, respecting the value of human life requires responding appropriately to the rational nature of this life. It requires that we engage with other rational beings in such a way that recognizes their ability to assess reasons and guide their actions by them.¹²

The case of slavery is a good illustration. At least part of what makes the institution of slavery morally abhorrent is that it empowers some individuals to systematically deprive certain classes of persons of the ability to choose their own projects and pursue them. The slave-owners deprive the slaves of the ability to govern their lives on the basis of what reasons they take to be salient. Slavery is wrong because the slave-owner subverts the slave’s rational autonomy. But while intuitions about the wrongness of slavery are virtually unanimous, the same unanimity cannot be found in the case of treatment of animals. On the face of it, our moral standards about the ways in which it is permissible to treat animals seem less strict. While a law that would prevent individuals of a certain ethnicity from entering public buildings would be clearly wrong,

¹² Scanlon, *What We Owe to Each Other*, pp. 103-4.
it is less obviously wrong to have a policy according to which no pets are allowed in public buildings. At least part of the explanation for this difference is that animals are not entitled to the same degree and kind of respect as are humans, because they fail to satisfy the requirements of rational autonomy.

For this account to be persuasive, however, there must be a close and plausible connection between the respect owed to rational creatures and the rational creature’s capacity for rational self-governance. On Scanlon’s account:

...respecting the value of human (rational) life requires us to treat rational creatures only in ways that would be allowed by the principles they could not reasonably reject in so far as they, too, were seeking principles of mutual governance which other rational creatures could not reasonably reject.”

According to Scanlon, showing proper respect to rational life requires that we act in ways (or display attitudes) that recognize their rational autonomy, that is, their capacity to assess reasons. We do this just in case our actions are justifiable on the basis of principles of conduct. On this account, to wrong a rational being (i.e. to fail to show the appropriate respect for rational life) is to act in ways (or, more generally, to display attitudes) that fail to be justifiable to others on grounds they could not reasonably reject. The thought here is that, in so far as human beings are rational, they are capable of assessing reasons. Therefore they are capable of both justifying their conduct, and understanding the justifications other offer. These activities (i.e. providing and understanding justification) take place within the space of reasons and in light of what reasons there are.

What is important to note here is the appeal of standing in this relationships with

13 Scanlon, What We Owe to Each Other, p. 106.
others (a relationship in which one’s actions are not reasonably rejectable). Human beings have reasons to want to stand in this relationship to each other. These reasons are not merely prudential. Living with other rational beings on terms that no one can reasonably reject is something we have reason to pursue for its own sake. Moral agents find moral criticism troubling precisely because it claims that they have failed in maintaining this relationship. This invariably impairs the offender’s relationship with others. Such an agent can no longer interact with others as someone who can successfully answer demands to justify his conduct. This alienates the offender from others.

Therefore on Scanlon’s account the impairment of one’s relationship with others (because of one’s having shown attitudes that are reasonably rejectable) is both necessary and sufficient for the appropriateness of moral criticism. This impairment can arise only with respect to those states (for example beliefs, desires, intentions, etc.) for which it makes sense to demand and offer reasons and justification. Scanlon calls states with such a content judgment-sensitive attitudes:

As rational creatures, we are capable of making judgments about reasons and hence of having judgment-sensitive attitudes such as belief and intention. In calling these attitudes judgment-sensitive, I do not mean to suggest that they always arise from conscious judgment. My point is, rather, that it is part of the nature of such attitudes that, in so far as we are rational, we come to have them when we judge ourselves to have compelling reason of the relevant kind to do so, and cease to have them when we judge there to be compelling reason against them.  

Attitudes can be judgment-sensitive in either of two ways. In the first place, they may arise from a judgment of the agent about the force of reasons. But while many of our attitudes do arise out of a conscious judgment about the force of reason, that is, out of a 

conscious decision to consider one set of considerations as decisively counting in favor of some attitude, on Scanlon’s account for an attitude to be judgment-sensitive and therefore appropriately subject to moral criticism, it need not have arisen in this manner. Judgment-sensitivity ensures that there exists cognitive harmony between various elements of the agent’s psychology, and in particular between the agent’s attitudes and his considered judgments.\textsuperscript{15} This cognitive harmony consists in the fact that judgment-sensitive attitudes are naturally sensitive to and fall in line with one’s conclusions about the force of reasons. The point here is that in so far as human beings are rational, they are capable of attitudes with a particular sort of content (i.e. that certain considerations count in favor of having certain attitudes). Given this particular sort of content, these attitudes are naturally sensitive to one’s conclusions about reasons. If one concludes that an attitude is not warranted, then one’s attitudes are ideally bound to respond to this conclusion. In this case, one no longer would be disposed to manifest the attitude in question. What is important to note here is that while a discrete judgment that a certain set of considerations count in favor of certain attitudes is sufficient for the judgment-sensitivity of that attitude, it is not necessary.

Consider the example of someone who is born into an extremely racist family. Since very young age, he is taught that individuals of a certain ethnicity are inferior and ought to be treated as such, and he associates only with individuals who hold such a view. Quite predictably, he comes to share these views. Such a person takes the fact of someone’s ethnicity to decisively count in favor of treating him in certain ways, for example denying him certain opportunities in life, not rendering assistance when he can

\textsuperscript{15} Scanlon, \textit{What We Owe to Each Other}, p. 21.
do so at little or no cost to himself, and perhaps even failing to uphold duties of regard and care otherwise owed to them. Suppose upon entering college, this person is exposed to certain arguments about the inherent worth of persons. Such theoretical considerations are supplemented and reinforced by the inevitable broadening of his social network, as a result of which he comes to be exposed to a variety of individuals of different ethnicities. As a result, this person no longer is inclined to count the fact of someone’s ethnicity as counting in favor of treating him in certain ways. Indeed, the fact of someone’s ethnicity now strikes this person as an altogether wrong sort of consideration to bear on questions about the moral standing of persons. What is important to note here is that in order for this conversion to take place, there need not be a discreet event of judgment that certain reasons do or do not count in favor of certain ways of acting. What is required is that the agent’s attitudes be sensitive to what one takes to be good reasons, which can be the case only if the attitude in question has the right intentional content.

Two points about the notion of judgment-sensitivity bear emphasis. The first is that the fact that an attitude is judgment-sensitive does not guarantee that one’s attitude do in fact track reasons. From the fact that an attitude is judgment-sensitive it does not follow that one is necessarily correct about what reasons count in favor of what attitudes. The agent in our example above, for instance, might have been utterly unmoved by arguments about the inherent worth and equality of persons. This would be a mistake on the part of the agent, but would not undermine the judgment-sensitivity of his attitudes, since whether or not an attitude is judgment-sensitive is, as I have argued above, primarily a question about its intentional content. Even if the agent is mistaken about
what reasons are decisive, it is nevertheless the case that he takes certain considerations to count decisively in favor of his attitudes.

Furthermore (and this is the second point) it is important to note the idealized nature of judgment-sensitivity. According to Scanlon, judgment-sensitive attitudes:

...are attitudes that an ideally rational person would come to have whenever the person judged there to be sufficient reasons for them and that would, in an ideally rational person, “extinguish” when that person judged them not to be supported by reasons of the appropriate kind.16

According to Scanlon, rational agents are agents capable of judgment-sensitive attitudes, capable of appreciating the significance of reasons and governing their behavior accordingly.17 This requires systematic, and not merely accidental, connections between various components of the agent’s psychology.18 Therefore, Scanlon understands irrationality narrowly, as severing the connection between these various elements of the agent’s psychology:

Irrationality in the clearest sense occurs when a person’s attitudes fail to conform to his or her own judgments: when, for example, a person continues to believe something (continues to regard it with conviction and to take it as a premise in subsequent reasoning) even though he or she judges there to be good reason for rejecting it, or when a person fails to form and act on an intention to do something even though he or she judges there to be overwhelmingly good reason to do it.19

Thus the cognitive harmony which is a distinctive feature of agents who are capable of judgment-sensitive attitudes can be thwarted by mental illness, or weakness of will.

16 Scanlon, What We Owe to Each Other, p. 20.
17 Scanlon, op. cit., p. 23.
18 Scanlon, op. cit., p. 25.
19 Scanlon, What We Owe to Each Other, p. 25; see also T. M. Scanlon, “Structural Irrationality,” in Geoffrey Brennan, et. al., Common Minds: Themes from the Philosophy of Philip Pettit (Oxford University Press, 2007) pp. 84-103.
Having taken an advanced course in philosophy of religion, for example, one may come to believe that none of the arguments offered in favor of the existence of god, understood as a supreme being that is omniscient, omnipotent and wholly good, is capable of establishing the conclusion in question. But one may nevertheless continue to passionately believe in such a being in order to fend off loneliness, or out of an instinctive need to secure an “ultimate purpose” for life. While such a person lacks the cognitive harmony that a fully rational agent would possess, the attitudes he displays are still judgment-sensitive and an appropriate basis for his moral criticism.\(^{20}\)

What is crucial to note at this point is that whether an agent recognizes the value of others as rational beings is centrally a question about the moral status that others enjoy from his point of view. We care about the agent’s attitudes only in so far as they reveal the force reasons seem to the agent to have, and in particular whether the agent understands and acknowledges reasons that flow from the value of human (i.e. rational) life. Since one’s attitudes are sensitive to judgment, they are susceptible to the force reasons seem to one to have. What reasons exist determine what attitudes are in fact justifiable, and not what attitudes the agent adopts in light of these reasons. The latter (i.e. what attitudes the agent chooses to adopt) depends crucially on what attitudes the agent takes to be justifiable, that is, the force reasons \textit{seem} to the agent to have.\(^{21}\)

Consider again the example of a racist who considers individuals of a certain skin color to be inferior. Such a belief is morally abhorrent, but nevertheless if he is pressed to justify his belief, he will cite a range of facts and considerations. According to him, these

\begin{itemize}
\item \(^{20}\) Scanlon, \textit{What We Owe to Each Other}, p. 20.
\item \(^{21}\) See, for example, Smith, “Control, Responsibility, and Moral Assessment,” p. 370. According to what Smith calls the ‘rational relations view,’ agents are answerable for their recognition of reasons that exist.
\end{itemize}
considerations count in favor of (i.e. justify) his attitudes. The fact that he takes these
facts to justify his attitudes, however, indicates these facts seem to him to enjoy a
particular salience. According to the racist, these considerations are not only, in the first
instance, relevant to the question of the moral standing of others, but also, in the final
analysis, decisive. But notice that this is possible only if the agent has a point of view
from which he is capable of considering various reasons for action and taking one or
more of them to be decisive. Our moral criticism of him, then, targets his point of view,
and in particular whether considerations he takes to be relevant to the moral standing of
others are indeed relevant, and, in the event they are, whether he assigns them the correct
importance. This, in turn, is important because the agent’s judgment-sensitive attitudes
reveal something about what the agent is like. It illuminates the quality of the agent’s
rational self-governance.

What is important to note is that on Scanlon’s account, whether an agent is an
appropriate candidate for moral criticism cannot be settled by any exclusively formal
condition. This manifests itself in two ways. In the first place, even if it is true that a
wholly mechanistic account of the agent’s psychology can be offered, this would not alter
what is most rational to do, that is, it would not alter what judgment-sensitive attitudes
are reasonably rejectable. Furthermore, and as a result, the truth of the Causal Thesis
does not have any bearing on whether the agent’s actions are a violation of the
requirements of good will, and whether his relation with others is impaired. This can be a
function only of the agent’s attitudes towards others, and in particular whether the agent
acknowledges reasons that flow from the value of others as rational beings by showing
appropriate regard for the well-being and legitimate interests of others.

This is an involuntarist account of preconditions of moral criticism. What is important to note is that, on Scanlon’s account, it is the fact that one’s attitudes have, in virtue of their content, implications for one’s relations with others, and not contra-causal freedom, or any claims about the agent’s abilities and powers, that makes the agent an appropriate candidate for moral criticism.

5.3 Voluntarist Criticisms of Responsibility as Attributability

I have argued in Section 5.2 that on Scanlon’s involuntarist account of preconditions of moral criticism, being an appropriate candidate for moral criticism does not require the ability of the agent to change his attitudes in light of reasons. This is because given their content, judgment-sensitive attitudes (such as belief, desire, intention, etc.) reflect the agent’s assessment of the reason-giving force of various considerations. It is this assessment that is the target of moral criticism.

Some philosophers have argued that Scanlon fails to successfully divorce moral criticism from the requirement of freedom and autonomy. R. Jay Wallace, for example, has focused on the case of wayward desires (desires that resist the agent’s attempts to modify or withdraw them) and argues that moral criticism of the agent for such attitudes is inappropriate:

...wayward desires...persist even in the face of conflicting reflective judgments about what there is reason to do, and require to be struggled against. Though wayward desires of this kind have a conceptual structure, their persistence in the presence of conflicting normative judgments suggests that they are not in fact open to direct rational control by the agent; this in turn makes the request to withdraw or modify the attitude latent in moral criticism seem misdirected.22

Wallace’s criticism of Scanlon rests on two premises. First, Wallace notes that on Scanlon’s account, moral criticism is constituted by demands addressed to the agent that he should revise or modify the attitudes that we take to be faulty. But, and this is the second premise, according to Wallace this demand is appropriate only if the agent is capable of revising his attitudes in accordance with his reflective judgment about the force of reasons. The demand for the withdrawal or modification of the faulty judgment-sensitive attitudes (which is implicit in moral criticism) makes sense only if the agent to whom this demand is addressed is not only capable of understanding what this demand consists in, but is also actually capable of complying with the demand. According to Wallace, the problem with wayward desires is that, *ex hypothesi*, this demand cannot fulfilled. This leads Wallace to conclude that having an intentional content is not a sufficient condition for responsibility as attributability, and therefore, that Scanlon fails to divorce responsibility as attributability from freedom.

The immediate question that arises out of Wallace’s criticism is whether he is correct in claiming that moral criticism of the agent for his recalcitrant attitudes is inappropriate and misplaced, and that moral criticism of the agent is appropriate only if the agent has the power to modify or withdraw them. Wallace’s claim is plausible only on the assumption that the problem of responsibility as attributability, and the question of what states an agent can be morally criticized for, is primarily a question about freedom, and in particular about the metaphysical bases of free action. He claims:

We blame people in the first instance for their decisions and choices, and for their actions to the extent that these reflect questionable decisions or choices. The presupposition of this practice is that states of these kinds are
subject to the direct control of the agent, in a way that mere desires do not necessarily seem to be; they are central examples of the class of volitional attitudes...[that serve] as the locus of self-determining agency. A moral psychology that leaves room for attitudes of this kind will thus have a principled basis for distinguishing between the forms of moral assessment appropriate to mere desires and emotions, on the one hand, and to actions and decisions on the other.23

According to Wallace, the question of whether some state is attributable to the agent (in the sense of being an appropriate basis for moral criticism of him) is primarily a question about whether the agent has “direct rational control” over that state, that is, whether he can exercise power (that is, freedom) over that state. On Wallace’s account, while decisions and emotions are similar in that they both have intentional content, it is only decisions (and actions that flow from them) that are necessarily under the control of the agent. Emotions, and in particular wayward emotions, therefore, are not appropriate bases for moral criticism of the agent, a conclusion which Scanlon’s account of responsibility as judgment-sensitivity fails to accommodate.

Similarly, in a recent paper John Fischer has argued that Scanlon fails to successfully divorce responsibility as attributability from freedom.24 Fischer’s main criticism of Scanlon is that Scanlon fails to offer a convincing case that responsibility as attributability depends on judgment-sensitivity:

The main objection I have to Scanlon’s approach is his linking responsibility as attributability with judgment-sensitivity; more specifically, the objection is to the suggestion that the basis for such moral responsibility is to be found in judgment-sensitive attitudes.25

25 Fischer, op. cit., p. 213.
Fischer’s argument for the conclusion that responsibility does not require judgment-sensitivity proceeds in two stages. The first consists of the claim that judgment-sensitivity is to be understood counterfactually. On Fischer’s account, an attitude is judgment-sensitive only if, if the agent judged that it should be revised, he would be capable of doing so. Fischer holds that “judgment-sensitivity (with respect to Y) would be present only if the following two conditionals would be true: “‘If the individual were to judge Y best, he would choose to do Y,’ and ‘If the individual were to judge not-Y best, he would choose and do not-Y.’”26 The thought here is that an attitude is actually judgment-sensitive only if it (i.e. the attitude) is revisable should the agent judge that it is unwarranted.

The second premise holds that judgment-sensitivity (counterfactually understood) is not necessary for blameworthiness. Fischer here deploys a Frankfurtian example,27 with the aim of showing that the ability to actually modify one’s attitudes (once one judges that doing so is warranted) is not necessary for one actually being blameworthy for the attitude one does hold. In Fischer’s example,

...a Frankfurt-case “counterfactual intervener,” such as Black, could render it true that a particular attitude is not judgment-sensitive without intervening in the scenario at all.... [he] could use the relevant judgment as a triggering event. He could thus block the connection between an

26 Fischer, “Responsibility and the Kinds of Freedom,” p. 220. It must be borne in mind that an agent can not do what he judges there to be decisive reasons for, but only on the pain of irrationality.
27 See Harry Frankfurt, “Alternate Possibilities and Moral Responsibility,” in Journal of Philosophy 66 (1969) pp. 829-39. In this essay, Frankfurt combats the principle of alternate possibilities, according to which the ability to do otherwise than one does is a necessary condition for the appropriateness of moral criticism. Frankfurt argues that an agent (Jones) may be unable to do otherwise than he or she actually does in virtue of the presence of a counter-factual intervener (Black). Black wants Jones to perform a certain action, and would intervene if Jones chose a different course of action. However, Black is never forced to intervene, because Jones decides on his own to perform the action Black wanted him to perform. Frankfurt takes this to be a counterexample to the principle of alternate possibilities. Jones can be appropriately held accountable even if he could not do otherwise, because he did what he did voluntarily.
alternative judgment and different attitudes and behavior.\(^{28}\)

This, however, does not render the agent non-blameworthy for the actual attitude he does hold:

In a suitably revised Frankfurt-case, the agent (say Jones) makes a judgment as to what is best based on his own reasons, and he is in no way impaired or interfered with. Further, this judgment issues in an appropriate attitude and also subsequent behavior. Intuitively, Jones acts freely and is morally responsible. But given the presence of [the counterfactual intervenor] Black, it is true that had Jones judged differently, Black would have swung into action and induced the very same attitude and behavior that occur in the actual sequence of events. Thus, Jones’s attitude is not judgment-sensitive...and his behavior does not flow from a judgment-sensitive attitude.\(^{29}\)

Fischer’s criticism requires careful analysis. On Fischer’s account, it is the fact of the presence of Black, and not anything that Black actually does, that renders the attitudes in question non-judgment-sensitive. The same point would presumably hold in the case of the truth of mechanism. That is to say, if mechanism is true, it would follow that we are no more judgment-sensitive than animals and natural objects. Furthermore, and relatedly, it is important to emphasize that on Fischer’s account, the presence of the counterfactual intervenor does not undermine the appropriateness of moral criticism of the agent. Moral criticism is not what is at stake here. Rather, Fischer’s point is that the agent continues to be an appropriate candidate for blame in spite of the fact that his attitudes do not satisfy conditions of judgment-sensitivity. This, according to Fischer, shows that judgment-sensitivity is not a necessary condition of moral criticism.

Fischer’s criticism, just like that of Wallace, hinges on the claim that the problem of responsibility as attributability is centrally a problem about freedom, and in particular

\(^{28}\) Fischer, op. cit., pp. 213-4, emphasis in the original.  
about freedom of action. On both accounts, judgment-sensitivity can be thwarted by the absence of the requisite element of freedom. While Wallace is concerned with how this freedom (and therefore judgment-sensitivity) can be undermined by facts about the agent’s psychology, Fischer focuses on impediments external to the agent’s psychology. Nevertheless, both Wallace and Fischer suggest that agents who fail to satisfy the freedom-relevant condition are inappropriate candidates for moral criticism, even if they satisfy Scanlon’s normative account of conditions of moral criticism.

In the case of Wallace in particular, there are two ways in which this inappropriateness is manifested. Wallace’s criticism can be understood as claim about the accuracy of morally criticizing an agent for a wayward attitude. Recall that in the case of wayward desires, the agent concludes that there are sufficient reasons to modify or withdraw an attitude, but his attitudes fail to respond to this conclusion. On this understanding, Wallace claims that in so far as we care about agents’ rational agency, to morally criticize the agent for an attitude that defies his assessment of reasons is to misidentify the agent’s attitudes, that is, to criticize the agent for an attitude that is not his. Alternatively, Wallace can be understood as making a claim about the appropriateness of making a demand of the agent that he simply cannot fulfill. On this account, since the agent cannot, ex hypothesi, revise or modify his attitudes, to demand that he should do so is to place a burden on him that is excessive and unwarranted. This criticism is particularly poignant in cases in which the agent readily acknowledges that his attitudes are faulty, but is, as a matter of fact, helpless with respect to them. To morally criticize him for such attitudes seems to impose an overly burdensome cost on him.
Therefore, Wallace’s and Fischer’s criticisms give rise to the question of whether it is the case that moral criticism is primarily a question about free action. There are two subsidiary questions here, namely (1) whether moral criticism is rendered inappropriate by the costs entailed by moral criticism, and (2) whether it is the case that to morally criticize an agent for states that have the requisite intentional content but defy the agent’s control is to misidentify the target of moral criticism. I will consider each question in turn.

5.4 Choice and the Burdens of Moral Criticism

As I have argued in Section 5.3, one intuition behind Wallace’s and Fischer’s criticism of Scanlon’s involuntarist account of moral criticism is a worry about the cost moral criticism imposes. Moral criticism is not pleasant. The claim that the agent’s attitudes are faulty is a claim about what he is like as an agent. Furthermore, the demands implicit in moral criticism, namely for justification of the attitudes or their modification, are themselves costly. This is particularly true in cases with which Wallace is concerned. In such cases, the agent’s faulty attitudes are inconsistent with his assessment of reasons. The thought here is that to morally criticize the agent for such attitudes, that is, to demand that he should withdraw his faulty attitudes, is to impose a requirement on him that he simply cannot satisfy.

As I noted in Section 5.2, however, a central feature of Scanlon’s account of moral responsibility is a distinction between two ways in which ascriptions of moral responsibility can be employed. On his account, questions of responsibility as attributability are concerned with whether the agent satisfies preconditions of moral
criticism. Nothing is implied about what sorts of reactions are appropriate, but only whether the attitudes the agent exhibits are his, such that it would be appropriate to take them as a basis of morally criticizing him.\textsuperscript{30}

To make this clear, consider the distinction between cases in which moral criticism is rendered altogether inappropriate and those that it is not. Features like height and eye color do not attract moral criticism, because they lack the requisite intentional content. By contrast, one may be morally criticized for one’s moral failings, that is, for acting in ways that do not uphold the value of others as rational beings and are therefore reasonably rejectable. But from the fact that one may be morally criticized for such features nothing follows about precisely what sorts of responses are appropriate. This latter question is a question of substantive responsibility, that is, a question about the sort of burdens we can appropriately require an agent to bear as a result of his actions.

At least intuitively, not all actions that are appropriate bases for moral criticism warrant the same sort of reaction. Consider the case of someone who exhibits racist attitudes as a result of being born into a racist family and having undergone a process of moral education that instilled in him such values. Compare this case with that of someone who is born into a good family, and has had the advantage of a progressive education, but who, nevertheless, ends up having racist attitudes. Intuitively, there seems to exist a significant difference between these two individuals. However, the difference does not consist in the fact that only the latter individual is an appropriate candidate for moral criticism. This is because in both cases the agent’s action reveal something about him as an agent, and in particular that his attitudes fail to uphold reasons that flow from the

\footnote{Scanlon, \textit{What We Owe to Each Other}, p. 285}
value of humans as rational beings, and therefore reveal an impairment in his relationships. Instead, the difference between the two cases consists in the fact that the circumstances of each case renders appropriate a different way of engagement with the agent.

According to Scanlon, this difference can be accounted for by appealing to the value of choice account (VCA). This account is motivated by the idea that all agents have reasons for wanting the outcome of situations to depend on their choice. On Scanlon’s account choice has instrumental, representative and symbolic value. The significance that agents attach to having a choice diminishes the grounds they have for rejecting principles imposing burdens on them if they have been provided with a choice. According to VCA, moral significance ought to be attached to the agent’s responses to the options he is presented with, and what has moral significance is the fact that an agent could have avoided certain burdens by having chosen appropriately in light of his circumstances.

It is significant to note that while VCA does employ the notion of choice (that is, intention, or control) the role it assigns to it is importantly different from the role voluntarist accounts of responsible agency assign to choice in two ways. On voluntarist accounts, such as the forfeiture view, what matters in assessing whether an agent can be morally criticized for an outcome is whether the outcome in question can be traced to a choice of the agent. What matters here is whether the agent actually chose to bring about

31 Scanlon, What We Owe to Each Other, pp. 251-67.
33 Scanlon, op. cit., pp. 254-5.
34 Scanlon, op. cit., pp. 254-5.
35 Scanlon, op. cit., p. 256.
36 Ibid.
a certain outcome, or, alternatively, whether he chose some course of action which could reasonably be foreseen to bring about the outcome. On VCA, by contrast, it is not the fact of the agent’s choice that has moral relevance, but rather whether the agent has had the opportunity to avoid a given burden by having chosen appropriately. It is important to emphasize that whether an agent has had an adequate opportunity to avoid a burden is not a question about whether the agent actually chose his current circumstances, nor a matter of availability of metaphysically open alternate possibilities. Instead, what is required is that the agent be put in good enough a position, such that he can avoid burdens by choosing appropriately.

It follows that the claim that moral criticism of agents for their wayward desires is rendered inappropriate because of the unavoidable costs imposed by moral criticism is false for two reasons. In the first place, the question of whether it is appropriate to require of an agent that he should bear a certain cost is governed not by considerations about responsibility as attributability, but rather by questions of substantive responsibility. The fact that a certain desire is impervious to the agent’s control, and that therefore the demands imposed by moral criticism cannot be met, does not entail that moral criticism of this agent is inappropriate. What is required here is an account of the agent’s substantive responsibilities. In the second place, the agent’s substantive responsibilities, the burdens he can legitimately be required to bear (in the form of the sorts of reactions his wayward desires render appropriate) does not depend on any claim about the agent’s freedom to do otherwise, but rather only on whether the agent was put in such

circumstances that he could have avoided burdens by responding appropriately to his circumstances.

Let us briefly consider the example of the racist who displays the attitudes in question because of being born into a racist family, and being surrounded by and educated in a community who accepts and propagates such views. The fact that such a person’s attitudes can be explained by the fact of his family circumstances and education does not render the heavy burdens associate with moral criticism (of isolation, alienation, censure, ostracism, etc.), inappropriate. This is because the facts that explain his attitudes are incapable of changing the fact that he still does display faulty attitudes (attitudes that are reasonably rejectable). Similarly, facts about his formative circumstances are incapable of altering the fact that such an agent’s relationship with others is impaired. The story about his formative circumstances, however, does suggest that he has been deprived of the opportunity to develop his character in an appropriate manner. This not only mitigates our response towards him, but also suggests that we have certain obligations towards this individual. We may owe him a progressive re-education, an education on the basis of a correct account of the worth of persons. More importantly, we may owe the community of which the agent in question is a representative. We owe this community the removal of barriers to knowledge. We owe them the adoption of measures that would extricate them from their ignorance. What is crucial is that while the fact of his history constitutes neither an excuse nor an exemption, it nevertheless does constrain the burdens we can appropriately require him to bear, and it may, in fact, place burdens on the rest of us to ensure that the agent in question is placed in an appropriate situations where he can
avail himself of the instrumental, representative, and symbolic value of choice.

In this section I have rejected the claim that the costs associated with moral criticism are rendered inappropriate in the case of wayward desires. In Section 5.5 I turn my attention to the question of whether moral criticism of the agent for wayward desires is unfair on the ground of inaccuracy.

5.5 Responsibility as Attributability and Ability to do Otherwise

In Section 5.4 I argued that concerns about the appropriateness of imposition of burdens associated with moral criticism for wayward desires is unwarranted. I argued that the question of appropriateness of moral criticism is a question about responsibility as attributability, and not substantive responsibility. Considerations about choice and control figure into our account of substantive responsibility, and not whether moral criticism of agents is appropriate. In this section I turn to the question of whether moral criticism for wayward desires is inappropriate on grounds of inaccuracy of blaming someone for holding an attitude that he or she cannot alter. I argue that paying attention to the structure of judgment-sensitive attitudes diminishes the plausibility of Wallace’s and Fischer’s criticisms.

Begin by considering Fischer’s claim that the presence of a counterfactual intervener is sufficient to undermine judgment-sensitivity. We can begin by noting that an attitude always has some content, namely that such and such a reason counts in favor of such and such an attitude. Neither Fischer nor Wallace disputes the claim that this is at least a necessary condition for the appropriateness of moral criticism. Therefore, for the counterfactual intervener to be able to render a certain attitude non-judgment-sensitive,
he must be capable of either of two things. He must be capable of altering the relevant intentional context, and make it the case that no reason can count in favor of (or against) the attitude in question. As a result, he must be capable of making it the case that reasons do not have a seeming force to the agent, and that the agent is incapable of assessing the force of reasons. Alternatively, he must actually intervene and prevent the agent from acting on the basis of the reasons he (i.e. the agent) takes to count in favor of the attitude in question.

But even on Fischer’s own formulation, the counterfactual intervener enjoys neither of these two powers. For on Fischer’s own formulation the counterfactual intervener lacks the power to make it the case that reasons lose their normative status. Indeed, it is difficult to see how any agent could have such a power. To count in favor of attitudes is just what being a reason is. Thus unless one thinks that the mere fact of the presence of a counterfactual intervener entails skepticism about the very existence of reasons, it is not clear whether the presence of a counterfactual intervener can render an attitude non-reason-responsive. Furthermore, since the Frankfurtian intervener is a *counterfactual* intervener, he never intervenes in the actual sequence. Therefore, he can never make it the case that the agent actually acts against his assessment of the force of reasons. Therefore, the mere presence of a counterfactual intervener does not render the attitude in question non-judgment-sensitive.

Since considerations about action and Frankfurtian scenarios are often accompanied by metaphysical baggage, one way to fully appreciate the force of the argument I have just offered is by considering beliefs.\footnote{Cf. Pamela Hieronymi, “Controlling Attitudes,” in *Pacific Philosophical Quarterly* 87 (2006) pp. 45–74.} Clearly, rational agents take
certain reasons to count in favor of holding certain propositions to be true. Thus if we ask such a rational agent why he takes a given proposition to be true, he will offer a range of considerations that he takes to count in favor of holding the proposition in question to be true. But it is not clear that the presence of a counterfactual intervener can by itself alter this fact. Even in the presence of a counterfactual intervener there are considerations that count in favor of holding certain propositions to be true. By the same token, even in the presence of a counterfactual intervener the agent’s attitudes are informed by what he takes these reasons to be. For the counterfactual intervener to render an attitude non-judgment-sensitive, he must be capable of making it the case that no consideration counts in favor of any attitude, a power that he lacks.

If I am right in claiming that the mere presence of a counterfactual intervener does not undermine the judgment-sensitivity of an attitude, the presumption that judgment-sensitivity is primarily a question about free action becomes implausible. I think a similar flaw can be detected in Wallace’s account. Recall that on Wallace’s account, moral criticism of an agent for his wayward desires is inappropriate. This is because Wallace holds that implicit in moral criticism is a demand for the agent to modify or withdraw an attitude. But this demand seems inappropriate if the agent is, as a matter of psychological fact, incapable of doing so.

But judgment-sensitivity is not, contrary to what Wallace suggests, primarily a matter of demanding modifications to the agent’s faulty attitudes. Scanlon writes:

...moral criticism claims that an agent has governed himself in a way that would not be allowed by any principles that no one could reasonably reject. When addressed to a participant in a system of co-deliberation, this charge calls for her to explain why this claim is mistaken, or to
It should be clear that, on Scanlon’s account, judgment-sensitivity, as a precondition of moral criticism, is a question about the relationship that exists between the agent’s attitudes and reasons that flow from the value of persons as rational beings. As I have argued earlier, this relationship is constituted by the property of justifiability of the attitude to other agents on grounds they could not reasonably reject. Thus the primary aim of moral criticism is to demand justification for the judgment-sensitive attitude which are claimed to be faulty. Moral criticism claims that the agent has failed to recognize reasons that follow from the value of rational beings, and demands that the agent show that this claim is false (that the attitude in question does not flout reasons). This demand for justification metamorphoses into a demand for the withdrawal or modification of the attitude only if the moral dialogue it instantiates fails, that is, only if either no justification is forthcoming (i.e. the agent simply refuses to engage in moral dialogue) or else the agent’s justifications are not satisfactory in light of shared principles. In either case the conclusion that the attitude in question is faulty goes undefeated, and the demand for justification evolves into a demand for the withdrawal or modification of the faulty attitude.

It follows that moral criticism is inappropriate only if the demand for justification it embodies is inappropriate. But, as I have argued, the question of whether or not this demand is appropriate is primarily a question about the content of the state in question. For example, as I have noted earlier, the reason it makes no sense to demand justification for one’s height or country or birth is that such features lack the requisite intentional

---

39 Scanlon, *What We Owe to Each Other*, p. 268, emphasis added.
content, whereas judgment-sensitive attitudes do not. But, and this is the crucial point, whether or not the agent is capable of revising his judgment-sensitive attitudes has no bearing on whether or not these attitudes have the intentional content that makes them sensitive to reasons. As a result, the question of whether the agent possesses the freedom to modify or withdraw his faulty attitudes is orthogonal to the question of whether he is an appropriate candidate for moral criticism.

This conclusion is further confirmed by considering the question of conditions under which it is appropriate to suspend or modify moral criticism, as embodied by Strawsonian reactive attitudes. Our starting point was Strawson’s insight that agents are not normally indifferent to wrongdoing and harms inflicted. On Strawson’s account, this is centrally because wrongdoing and harm are taken to reveal an attitude on the part of the wrongdoer. Counterconditions to moral criticism function by showing that this is not the case, that is, by showing that the agent’s actions do not reveal the attitudes in question, or, even if they do, the agent should be exempted on grounds of moral incompetence. With this in mind, recall that on Scanlon’s account judgment-sensitivity is an idealized notion. This means that an agent’s attitudes, given their content, are typically responsive to one’s reflective conclusions about the force of reasons. This leads to a harmony between the various components of the agent’s psychological economy. However, this harmony can be disturbed by failures of rationality, that is, by cases in which the agent’s attitudes fail to respond to one’s assessment of reasons. The case of wayward desires is one such example. In such cases, while the agent concludes that there are decisive reasons to modify or withdraw his faulty attitudes, he is incapable of doing
so. Therefore, for freedom to enjoy the distinct normative force that it does on Fischer’s and Wallace’s accounts, that is, for it to be the case that the ability of the agent to revise or withdraw his faulty attitudes is a necessary condition of moral criticism, it must be the case that irrationality is itself an excuse. It must be the case that irrationality entails that the agent does not have the attitudes his actions suggest he does, or else that he is morally incompetent.

I want to focus on the claim that irrationality entails that the agent does not have the attitudes his actions suggest. The intuitive idea here is that since the agent’s attitudes defy his reflective assessment of reasons, it would be inaccurate to attribute the agent’s faulty attitudes to him. In so far we place an importance on rationality, it is his reflective assessment of reasons that should be attributed to him, and not his faulty attitudes (which he reflectively disowns). For this suggestion to be correct, it must be the case that the agent’s faulty attitudes, which, ex hypothesi, are inconsistent with his reflective assessment of reasons, do not reveal an attitude on his part about the moral standing of others. But this is implausible.40

To see this, begin by considering the case where the agent’s cognitive and executive faculties are functioning optimally. Such an agent’s faulty attitude is in accord with his considered judgment about the status of reasons. His attitude reveals what his assessment of the force of reasons is. The fact that his attitude is faulty is due to a mistake either about facts that obtain, or in his assessment of the relevance of facts in his deliberation. Had the agent reached a different assessment than he actually did, had he

judged that there are decisive reasons for modifying or withdrawing his attitudes, he would have succeeded in effecting such a modification in his attitudes. Such a person is clearly an appropriate candidate for moral criticism.

Consider, for example, the case of Joliet I discussed in Section 4.3:

*Home for the Holidays.*—Joliet, who is afraid of burglars, is alone in the house. Panicked by sounds of movement in her kitchen, she grabs her husband’s gun, tiptoes down the stairs, and shoots the intruder. It is her son, who has come home early for the holidays.\(^4\)

Joliet’s action of shooting the person who had entered the kitchen late at night and was the source of the noise in the kitchen was an intentional action. If she is asked why she fired her gun, she would be able to justify her act by reference to her fear of burglars, as well as a belief that there were intruders in the kitchen. However, her assumption that the commotion in the kitchen is attributable to burglars turned out to be mistaken. This mistake was not committed in spite of Joliet’s due diligence, but rather because she acted instinctively and failed to fully investigate the circumstances. Furthermore, she was arguably mistaken in judging that the best way to deal with the intruder, even if there were one, was to shoot him, though perhaps this claim depends on a variety of factors, a full analysis of which is beyond the scope of the present discussion. What I want to emphasize is that the fact that Joliet’s (faulty) attitude is attributable to these two mistakes does not make moral assessment of Joliet inappropriate. In fact, it is precisely because she commits these errors (both in terms of fact-gathering and in terms of reason-assessment) that we consider it appropriate to morally criticize her, although the fact that

her mistakes were in good faith may mitigate our moral criticism, altering the particular responses towards her we deem appropriate.

The same is true in the case of a wholehearted racist (call him racist\textsubscript{1}). Racist\textsubscript{1} takes the fact of someone’s ethnicity to count in favor of treating him as inferior. Had this agent made a different assessment about the relevance of the fact of someone’s ethnicity to his moral standing, he would have been capable of renouncing his racist attitudes. Such a person’s faulty attitudes, therefore, are not due to a defect in his psychology, but rather due to a mistake about reasons. But this fact does not make his faulty attitudes any less his and therefore open to moral criticism. The very fact that the agent has a tendency to overlook the relevance of certain considerations, or make mistakes about which of these considerations are decisive, tells us something relevant about his attitudes, about what he is like as a moral agent. Our reactions to his faulty attitudes are not rendered inappropriate simply because he tends to make mistakes about what reasons there are.

There exists a noteworthy analogy between on the one hand cases in which a faulty attitude is attributable to factual error (i.e. error in the assessment of what facts obtain, as in the case of Joliet) or normative error (i.e. error about what reasons there are, as in the case of racist\textsubscript{1}) and on the other cases in which a faulty attitude can be attributed to errors (or malfunctions) within the agent’s psychology. This is centrally because the fact of the agent’s irrationality is itself revealing about whether he acknowledges reasons that flow from the value of others as rational beings. Cases of irrationality are not problematic because in such cases the agent’s faulty attitude is devoid of intentional content, but rather because the content of the agent’s attitudes are inconsistent. In cases of
irrationality (cases in which the agent judges that there are sufficient reasons for modifying or withdrawing an attitude but is, as a result of a defect within his psychology, incapable of effecting such a change), the agent has two inconsistent judgment-sensitive attitudes. One such attitude is the one reflected in his assessment of reasons, according to which there are reasons which legislate in favor of withdrawing or modifying the attitude that he concludes to be faulty. At the same time, his faulty attitude itself reveals a set of moral commitments. The fact that the faulty attitude in question is impervious to the agent’s conclusions about the force of reasons indicates not that the attitude in question is non-judgment-sensitive, but rather that the attitude in question is too deeply rooted to be dislodged by the agent’s judgment about the force of reasons. Therefore the fact that the agent is incapable of renouncing a faulty attitude even if he concluded that doing so is warranted, itself reveals something significant about what the agent is like, and in particular about what attitudes, from his point view, are justifiable in light of reasons. Even if the agent lacks the power to alter his attitudes, this very fact impairs his relationship with others.

Consider the example of someone who continues to have racist urges contrary to his considered judgments (call him racist2). While he concludes that there are decisive reasons against considering individuals of a certain ethnicity to have an inferior moral standing, he continues to consider them to do so. According to the proponents of the voluntarist account of judgment-sensitivity, there is a morally relevant difference (as far as responsibility as attributability is concerned) between racist2 and racist1. According to these philosophers, this difference consists in the fact that while the demand to withdraw
or modify the faulty attitude implicit in moral criticism does make sense when addressed to the wholehearted racist, it is misplaced and inappropriate in the case of the conflicted racist. But to establish this it is not sufficient to show that the conflicted racist’s attitudes defy his attempt to revise them. A stronger claim is required, namely that the conflicted racist’s attitudes are not really his to begin with (i.e. is not attributable to him at all).

The latter claim, however, is difficult to make plausible. For while the predicament of racist$_2$ is not enviable, his recalcitrant racist attitudes still arise from his assessment of the relevant features of the situation. Even if he is incapable *articulating* the considerations which he takes to count in favor of his racist attitudes, it seems that it must nevertheless be the case that he takes certain considerations to count in favor of his attitudes. This is evidenced by the fact that racist$_2$ is not racist towards anyone whatsoever. Rather, his negative reactions are directed towards individuals with particular characteristics. This indicates that he takes these characteristics to be relevant to persons’ moral standing. Therefore, the problem of racist$_2$ is not that his racist attitudes cannot be attributed to him, but rather that there are two contradictory attitudes that can be. This, however, does not show him to be an inappropriate candidate for moral criticism with regard to either of the two of his attitudes.$^{42}$

I have argued that irrationality (that is, the dissonance between the agent’s assessment of reasons and his attitudes) does not undermine the attributability of the attitude to the agent. This claim can be further substantiated by considering a third permutation of the example of the racist individual (call this racist$_3$). In this case, the

agent begins as a normally morally upright person who is hypnotized to react negatively
to individuals of a certain skin color. It seems overwhelmingly plausible to claim that that
racist3 is not an appropriate candidate for moral criticism for the induced racist attitude. It
might be tempting to observe that this agent is incapable of actually revising his attitudes
if he judged doing so to be warranted, and take this to be a basis for the conclusion that
he is not an appropriate candidate for moral criticism. This is because the hypnosis, if
effective, effectively circumvents the agent’s own rational capacities. But this
observation, even if correct, is not adequate, because the agent’s inability to actually
modify or withdraw his attitudes is not peculiar to this third case, but is rather shared by
racist2. At the same time, it seems clear that there is a crucial difference (as far as the
agent’s responsibility as attributability is concerned) between on the one hand racist3, and
on the other racist1 and racist2, in so far as the notions of judgment-sensitivity and
responsibility as attributability are concerned.

I want to highlight, more specifically, two such differences. Note, in the first
place, that racist3’s attitude lacks intentional content. The crucial thought here is that his
attitudes do not reveal his assessment of the force of reasons. His actions (if they are
actions) do not reveal what facts, if any, he takes to count in favor of holding the attitudes
in question, nor do they disclose that, from his point of view, reasons that flow from the
value of persons as rational beings are not acknowledged. Therefore, in such a case, the
question of whether his attitudes are reasonably rejectable does not even arise, and the
fact of his actions under hypnosis do not impair his relationship with others. Agents such
as racist3 do not act from a morally faulty quality of will. In the second place, the fact that
racist, is, *ex hypothesi*, a normally morally decent person, in concert with the fact that his racist attitudes is the result of hypnosis, means that his faulty attitudes are “just visiting,” in the sense that they would disappear once the hypnosis ceases to be effective.

Note that while while racist is incapable of actually modifying or withdrawing his faulty attitude, the attitude itself is informed by an assessment of the force of reasons, informed by considerations that the agent takes to count in favor of his attitudes. This assessment might be explicit, as in the case of the wholehearted racist, who would reflectively endorse his faulty attitude, or it might be implicit and in tension with other judgment-sensitive attitude within his psychological economy, as in the case of racist. But in either case, the agent’s attitudes are constituted by certain considerations that the agent takes, though mistakenly, to count in favor of his racist attitudes. Furthermore, the fact that their attitudes are so informed means that the faulty attitudes in question are not just visiting. For example, while racist’s racist attitudes will dissipate once the hypnosis wears off (on the assumption that he is normally a morally good person), in the case of neither racist nor racist is the recovery as automatic or easy. Therefore, racist is an inappropriate candidate for moral criticism not because he cannot, even if he tried, renounce his faulty attitudes, but rather because his implanted attitudes do not tell us anything about what he is like, and, more specifically, about what moral standing rational agents seem to have from his point of view (i.e. whether he acknowledges reasons that flow from the value of others as rational beings). And while racist does struggle against his faulty attitudes, he is not entitled to the excuse racist is.

In this section I have investigated the question of whether moral criticism of an
agent for a faulty attitude he is incapable of revising or withdrawing is inappropriate on grounds of inaccuracy. I have argued that whether or not moral criticism of an agent is inaccurate depends on whether the agent’s attitudes possess the content purported, and not whether he is capable of revising the attitudes. In cases Wallace has in mind (those in which the agent struggles against a faulty attitude but fails to dislodge it), it is not the case that the agent has only one attitude (one identified with his struggle against his faulty attitude). The correct way to understand such cases is that such agents in fact have two conflicting judgment-sensitive attitudes. While such agents may deserve praise for recognizing that a particular attitude of theirs is faulty and for struggling against it, this would not change the fact that they still hold the faulty attitude in question. They may be appropriately morally criticized for this. Therefore, I conclude that morally criticizing an agent for wayward desires is not inappropriate on grounds of inaccuracy.

5.6 Concluding Remarks

The question I investigate in this chapter is the plausibility of the requirement of up-to-us-ness, and the possibility of articulating a plausible non-freedom relevant condition on moral criticism. I argued in Section 5.2 that on Scanlon’s account, whether an agent is an appropriate candidate for moral criticism for a state is primarily a matter of the implications of the agent’s attitudes for his relationship with others, and in particular whether the agent’s attitudes are justifiable to others. In Section 5.3 I focused on Wallace’s objection to Scanlon’s association of responsibility as attributability with judgment-sensitivity. According to Wallace, the problem of responsibility is primarily a problem about free action. This freedom can be thwarted either on psychological grounds
(as in the case of Wallace’s wayward desires) or on metaphysical grounds (as in the case of Fischer’s counterfactual intervener). I argued that Wallace’s claim that appropriateness of moral criticism is primarily a matter of freedom is motivated by two thoughts. The first is that claiming an agent for his recalcitrant attitudes is to misidentify the target of moral criticism. The second is that to morally criticize an agent for recalcitrant attitudes is to impose an inappropriate burden on the agent. In Sections 5.4 and 5.5 I argued that both of these claims are false.

In Section 5.4 I argued the second criticism, namely the claim that to morally criticize the agent for an attitude that lies beyond his voluntary control is to impose an unreasonable burden on the agent, conflates two senses of responsibility Scanlon is at pains to distinguish. Whether an agent can be reasonably expected to bear a cost is a question about the agent’s substantive responsibilities. On this account an agent cannot reasonably complain about a cost he is required to bear if he was placed in a good situation such that he could have avoided the burden by having chosen appropriately. By contrast, the question of whether an agent is an appropriate candidate for moral criticism, in which I am interested in this chapter, is about responsibility as attributable and preconditions of moral criticism.

In Section 5.5 I argued that the second criticism is plausible only on the assumption that the inability of the agent to exercise control makes it the case that the agent’s attitude is no longer judgment-sensitive. But even if the agent is incapable of exercising control, this would not mean that the attitude in question is no longer one for which reasons can legitimately be demanded or offered. The fact that the agent is
incapable of bringing his attitudes in line with his assessment of reasons does not entail that the agent does not have any judgment-sensitive attitudes at all. Instead, it suggests that the agent holds two incompatible judgment-sensitive attitudes. This is illuminating as to what the agent is like, and therefore has implications for the agent’s relationships. Far from excusing the agent, his inability to bring his attitudes in line with his assessment of reasons is an appropriate basis for his moral criticism.

The argument of this chapter provides a basis for a normatively adequate interpretation of the quality of will thesis. For while I have argued in the previous chapters that the subjectivist and voluntarist interpretations of the quality of will thesis face significant difficulties, Scanlon is capable of offering a normatively rich interpretation of the quality of will thesis. On my reading, the significance of moral criticism rests on substantive normative considerations about reasons. Scanlon changes the focus of moral criticism from a sentiment of being wronged, that is, a feeling that the agent has failed to show appropriate regard for our well-being and interests, to reasons that flow from the value of humans as rational beings, and the significance of these reasons and their underlying values for the quality of interpersonal relationships.

The question of the relevance of control to the appropriateness of moral criticism therefore requires careful modification. It is the fact that we are agents capable of reasoning and having attitudes, and capable of interacting with agents who are likewise rational, capable of reasoning and having attitudes of their own, on which moral criticism is founded. The practices constitutive of moral criticism are not dislodged by the fact that we live in a natural world, a world governed by the same natural laws governing objects.
and things, because these laws themselves are not determinative of the rightness of the acts, nor can they decide whether any given relationship is impaired.
Bibliography


Recognition of Vocal Affect in Psychopathic Individuals.” *Journal of Abnormal Psychology* 111, pp. 682-86.


  *British Journal for the History of Philosophy* 15, pp. 27 – 57.

Debes, Remy. 2007. “Has Anything Changed? Hume's Theory of Association and
  Sympathy after the Treatise.” *British Journal for the History of Philosophy* 15, pp.
  313 – 38.


  Press.


Farrow, Tom F. D., and Peter W. R. Woodruff, eds. 2007. *Empathy in Mental Illness.*
  Cambridge: Cambridge University Press.

  and Criminal Responsibility: Psychopathy and the Purposes of Punishment.”


Haji, Ishtiyaque. “Psychopathy, Ethical Perception, and Moral Culpability.” Forthcoming in *Neuroethics*.


168


171


