Sentimentalism, Affective Response, and the Justification of Normative Moral Judgments

by

Kyle Martin Menken

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Master of Arts
in
Philosophy

Waterloo, Ontario, Canada, 2006

Author’s Declaration

I hereby declare that I am the sole author of this thesis. This is a true copy of the thesis, including any required final revisions, as accepted by my examiners.

I understand that my thesis may be made electronically available to the public.

Abstract

Sentimentalism as an ethical view makes a particular claim about moral judgment: to judge that something is right/wrong is to have a sentiment/emotion of approbation/disapprobation, or some kind of positive/negative feeling, toward that thing. However, several sentimentalists have argued that moral judgments involve not only having a specific kind of feelings or emotional responses, but judging that one would be *justified* in having that feeling or emotional response. In the literature, some authors have taken up the former position because the empirical data on moral judgment seems to suggest that justification is not a necessary prerequisite for making a moral judgment. Even if this is true, however, I argue that justifying moral judgments is still an important philosophic endeavour, and that developing an empirically constrained account of how a person might go about justifying his feelings/emotional responses as reasons for rendering (normative) moral judgments by using a coherentist method of justification is both plausible and desirable.

Acknowledgements

I would like to thank my supervisor, Patricia Marino, who, through her teaching, suggestions, and patience has made an immeasurable contribution not only to this essay, but to my philosophical development.

I would also like to thank my readers, Lorraine Besser-Jones and Paul Thagard, whose comments were always helpful, and who have helped show me the potential in a naturalistic approach to ethics.

Lastly, I would like to thank the Social Sciences and Humanities Research Council of Canada, the University of Waterloo, and the Department of Philosophy for their financial support over the last year.

Dedication

To my parents, Bernie and Irene, who have always stood behind me.

Table of Contents

Introduction

"Ethical Naturalism has yet to find a plausible synthesis of the empirical and the normative: the more it has given itself over to descriptive accounts of the origin of norms, the less has it retained recognizably normative force; the more it has undertaken to provide a recognizable basis for moral criticism of reconstruction, the less has it retained a firm connection with descriptive social or psychological theory."

Peter Railton, "Moral Realism."

As an ethical theory, sentimentalism bases morality in feelings, emotions, desires, passions, or sentiments. It has come under different guises in recent years, such as emotivism and expressivism. Despite the varying use of terminology, the defining trait is an emphasis on describing moral behaviour, judgment and knowledge in terms of what we feel as opposed to what we reason. Sentimentalists tell us that 'feelings' of one kind or another guide our moral lives; these feelings tell us what is wrong, what is right, and help us decide – or simply decide for us – what we morally ought to believe and do.

Sentimentalism makes a particular claim about moral judgment: to judge that something is right/wrong is to have a sentiment/emotion of approbation/disapprobation, or some kind of positive/negative feeling, toward that thing. This position on moral judgment marks out sentimentalism in its simplest from. However, many sentimentalists argue that such a simple sentimentalism fails to capture something distinctive and important about moral judgment. They claim that to judge that something is wrong is not just to have a negative feeling toward something, but rather to judge that it would be justifiable or appropriate to have some negative feeling toward that thing. I will refer to the former position as simple sentimentalism, and the latter as neo-sentimentalism. Proponents of simple sentimentalism include Shaun Nichols and Jesse Prinz; proponents of neo-sentimentalism include Alan Gibbard, Justin D'Arms, Daniel Jacobson, and, in

some of his moods, Simon Blackburn.[1]

Aside from their differences regarding moral judgment, simple sentimentalists and neo-sentimentalists differ in another important respect. In recent years, Shaun Nichols, and to a lesser extent Jesse Prinz, have capitalized on an emerging empirical literature that carves out a very important role for the *emotions* in moral behaviour and thought. Because of their emphasis on psychology and neuroscience, these authors emphasize developing a descriptive, rather than a normative, moral theory. Indeed, this fact is at the root of the disagreement between simple and neo-sentimentalists about moral judgment: neo-sentimentalism does not fare very well as a descriptive theory of moral judgment, whereas simple sentimentalism does not capture the normative aspects of moral judgment very well.

One hopes that it is at least possible that these differences can be reconciled. Indeed, the aim of this paper will be to show that a sentimentalist theory of moral judgment can be descriptively and empirically accurate while still maintaining a normative dimension, which in my opinion is necessary for it to be an acceptable *philosophical* moral theory. Because the difference between the opposing positions regarding moral judgment boils down to whether or not they believe that a person's emotional responses, or tendencies to have certain emotional responses, need to be justified in some important sense if that person is going to base (normative) moral judgments on those responses, the goal of this paper will be to outline how a person might go about justifying his emotional responses as reasons for rendering (normative) moral judgments.

---

[1] For a more detailed description of the defining characteristics of neo-sentimentalism, see Justin D'Arms and Daniel Jacobson's "Sentiment and Value." For more on simple sentimentalism, see Shaun Nichols, "Sentimental Rules," and Jesse Prinz, “The Emotional Basis of Moral Judgments.”

Before giving a brief outline of how this paper will proceed, a terminological issue needs addressing. In the following, instead of talking about people's emotional responses to things, I will talk about their 'affective' responses. The main reason for this is that I want to keep clear of the conceptual baggage that goes along with the word 'emotion.' For, as we will see, a person could have an affective response to something without it being true that she is in the grips of a particularly emotional experience, as most people understand 'emotional.' Moreover, people tend to think that if something is emotional then it cannot be cognitive, in the same sense as desires and beliefs are often taken to be distinct. However, this belief, as I will argue, oversimplifies what is at issue (see below, ch. 2, section 3). In addition, in the relevant psychological and neuroscientific literature one finds people using terms like 'affective response,' 'affect representation,' and 'affective tone' more often than one finds people using terms like 'emotional response,' 'emotional representation,' and 'emotional tone.' Based on these considerations, it seemed appropriate and natural to talk more so in terms of affect rather than in terms of emotion. With that out of the way, once more, the goal of this paper will be to outline how a person could go about justifying his *affective* responses as reasons for rendering (normative) moral judgment.

In chapter one, I will first examine simple sentimentalist theories of moral justification in some detail. This discussion will help carve out an important place for neo-sentimentalist construals of moral judgment in a more comprehensive sentimentalist moral theory. Second, through an examination of Shaun Nichols' argument against moral objectivity, I will show how affective responses are in a sense *prima facie* normative, just in virtue of their motivational force and phenomenological presence. This will lead into

the next section, which outlines three ways in which we could attempt to justify our affective responses: through self-justification, external justification, and internal justification, and argues that internal justification is the most promising method.

Chapter two tackles two salient objections to an emotion-heavy sentimentalist account of moral judgment put forward by Stephen Darwall, Alan Gibbard, and Peter Railton in their “Toward Fin de Siecle Ethics.” The first of these objections points to the fact that a person can make a moral judgment about something while having lost all disposition to feelings toward that thing, which leads one to think that feelings are not necessarily implicated in moral judgment. After replying to this objection, I go on to argue that people can indeed reason about their emotions, and make judgments on the basis of them, without being in what is commonly thought of as an emotional state. The second objection is more pointed. It questions whether 'feelings' or dispositions to have them are necessary for explaining moral judgments at all, if those feelings are in fact caused by a special kind of cognitive judgment in the first place, as many emotions theorists would have it. After replying to this objection, I argue that, explanations aside, a normative account of moral judgment needs to take into account the feelings we associate with affective response.

In chapter three, I propose that a suitably precise notion of coherence provides the best method whereby a person can justify his affective responses as reasons for rendering moral judgments. In the first section, I will analyze Linda Radzik's recent development of a coherentist account of normative authority, in order to formulate the conditions that a coherentist theory of the justification of affective responses should meet. The most important of these conditions will require that the theory be first-person accessible and

regress-avoiding. In the second section, I propose a formula for the justification of affective responses, which will rely on the *prima facie* normativity inherent in affective response. In fact, it is the normative quality of affective responses that will allow the theory to meet the first-person accessibility and regress-avoiding conditions.

Simple Sentimentalism, Moral Judgment, and Normativity

Despite their empirical sensibilities, simple sentimentalist theories often seem inadequate when applied to normative issues. Shaun Nichols has this to say about the normative status of our affective responses: "There is no principled basis for maintaining that these certain emotions (on which our moral judgments depend) are the *right* emotions." (Nichols 185). If we ask ourselves whether we really should believe or act on the basis of any affective response, I take Nichols to suggest, there are not any principled reasons on offer. As will become apparent, Nichols does not think that we need to justify our affective responses in order to take the moral judgments we base on them seriously. Rather, he thinks that we should accept our affective responses at face value.

For the sentimentalist concerned with the possibility of justifying affective responses, this position presents a pointed objection, and it deserves careful consideration. We need to examine why empirically informed sentimentalists take on this type of position, to see what its merits are, and to take what we can from it. To do so, I will describe how Nichols' theory of moral judgment is an attempt to develop a descriptive account of moral judgment, and that he succeeds in this task. Next, I will present Nichols' argument against moral objectivity, and in doing so I will criticize his notion of 'the persistence of moral judgment'. Following that, I will further analyze Nichols' views on the justification of affective responses, and I will use Simon Blackburn's exposition of a similar position to show both the pitfalls and the positives of the simple sentimentalist approach to justification of affective responses, and how these pitfalls can be avoided.

Nichols on Moral Judgment

Nichols' sentimentalist theory of moral judgment diverges from the neo-

sentimentalist view. Within it, for a judgment to be moral, two things must be implicated by that judgment: a normative theory, and an affective mechanism that is sensitive to suffering and distress of others (18). The normative theory he invokes "prohibits actions of a certain type, and actions of that type generate strong affective response" (18). If we think of the normative theory Nichols mentions as a body of rules governing behaviour, then moral judgments are simply invocations of rules that prohibit (or in some other way relate to) actions that cause people to suffer. Because of the connection between suffering in others and negative affect, these rules are infused with feeling, and hence Nichols calls them 'sentimental rules.' A neo-sentimentalist theory would require that a normative assessment about whether the response instantiated by the 'affective mechanism' is justified, appropriate, or warranted in some important sense. Nichols does not think that some kind of normative assessment must have taken place in order for a proper moral judgment to be made.

Why does Nichols reject the neo-sentimentalist approach to moral judgment? An important motivating belief is that neo-sentimentalist accounts of moral judgment are descriptively inaccurate (Nichols 89; Prinz 35). For, according to the standard neo-sentimentalist view, making a moral judgment involves making an assessment about the appropriateness of an affective response; it requires that people have the capacity for normative assessment if they are to be able to make genuine moral judgments. This approach is misled, Nichols says, because it seems that children and autistics, who lack the capacity for normative assessment as neo-sentimentalists describe it, make genuine moral judgments. Nichols terms this the disassociation argument: if people without the capacity for normative assessment do indeed make moral judgments, then we cannot say

that moral judgments necessarily have a normative element (89-90). Because children and autistics fit this description, we need to define 'core moral judgment' so that the moral judgments children and autistics make are included in that definition. Going further, he argues that many of the everyday moral judgments made by adults might not rely on some prior normative assessment (93-94).

This view makes some sense; many people are probably not as concerned about morality as ethicists would like them to be. It seems reasonable that a good description of moral judgment *qua* moral judgment should be broad enough to capture properly the judgments of children, autistics, and the more everyday moral judgments adults make. However, there still remains the fact that thoughtful, reflective people are concerned with the normative status of their moral judgments, and often take making moral judgments quite seriously. Notice that this fact leads to two distinct problems for simple sentimentalists. First, because of their emphasis on 'core moral judgment', they neglect to give a descriptive accounting of how people make normative assessments. Second, because of their focus on providing descriptive accounts in general, they have very little to say about how people should engage in normative assessment. Of course, Nichols is not necessarily insensitive to normative issues, and, as we will see, he makes some comments that help define his position in relation to certain normative problems.

Nichols is right that a properly descriptive theory of moral judgment should be more inclusive than the standard neo-sentimentalist construal. However there is an important difference between the moral judgments reflective people make and those made by children and people with autism. Non-autistic adults do have the capacity for normative assessment; many exercise it and consider the results of doing so important. Children's

moral judgments express what they feel and what have been taught. While thoughtful, reflective adults' moral judgments also often merely express what they feel and have been taught, the moral judgments made by thoughtful, reflective adults are different insofar as they are far more responsive to reasons, e.g., inconsistency with other judgments, or the distorting affects of situational influences.[2] While we need an account of moral judgment that is descriptively accurate and that allows people without the capacity for normative assessment to make genuine moral judgments, we can at least call moral judgments that *do* involve assessments of the appropriateness of feelings 'normative moral judgments'. Normative moral judgments, while perhaps a subspecies of moral judgment *simpliciter*, certainly have a life of their own, and deserve investigation both in the descriptive and the normative sense. Clearly, figuring out how justifying our affective responses gives rise to their being reasons, and how this process of justification should be responsive to certain kinds of reasons, is an important endeavour in its own right for the philosophical sentimentalist.

Objectivity and 'the Persistence of Moral Judgment'

One of the more ingenious elements in Nichols recent book "Sentimental Rules" is his analysis of moral objectivity. In that analysis, he uses the distinction between moral and conventional norms, which he takes from the psychological literature. The difference between moral and conventional norms (or rules, if you like) is best understood by the relation each bears to authority contingence. Using studies on this distinction that involve children, he writes: "For instance, if at another school the teacher has no rule against chewing gum, children will judge that it is not wrong to chew gum at that school; but

[2] Adults also grapple with more complex and weightier moral cases than children. Compare 'pulling hair is wrong' with 'genocide is wrong.'

even if the teacher at another school has no rule against hitting, children claim that it is still wrong to hit" (6). Moral violations, like hitting, are not contingent on authority, whereas conventional violations, like chewing gum, are contingent on authority. One can also understand this distinction by considering its relation to harm norms. Norms against harming others fall on the moral side of the distinction, whereas norms that do not tend to fall on the conventional side. Moral norms always implicate some kind of victim, where conventional norms do not (Blair et al. 15).[3] Further, he argues that harm norms have this status because they are keyed to the emotional distress that normal people feel when confronted by the suffering of others (64).

The relationship that Nichols posits between the moral/conventional distinction and objectivity is easily understood. He understands the basic tenet of moral objectivism to be that certain actions are wrong (right) non-relativistically, i.e., certain actions are wrong independently of individual preferences, beliefs, etc. (Nichols 169). If we consider the non-authority contingence, i.e. the non-relative status that people ascribe to norms that fall on the moral side of the moral/conventional distinction, we can see that moral norms seem to lay *prima facie* claim to being objective. Why do people think that moral norms are non-relativistic while conventional norms are relativistic? What difference is there between the two that can account for the difference in perceived relativity? In short, conventional norms are generally not infused with feeling by affective response that is caused by the suffering of others, whereas moral norms, and especially harm norms, are. Therefore, Nichols argues, it seems that "affect somehow infuses harm norms with objective purport" (179, 186).

[3] 'Disgusting' norms are an interesting exception. For instance, most people would say that it is always and everywhere wrong to spit on another person.

Now, if affect is what causes us to believe, pre-reflectively, that moral norms are objective, moral objectivity is on shaky ground. Unless we can show that a person rationally ought to have a certain affective repertoire, because that affective repertoire is somehow externally privileged (185), then morality is not in fact objective, but is relative to the affective responses each person (or group of people) has. Nichols argues that this is true by claiming that rational beings, such as aliens, for example, who simply did not manifest the same affective responses as us to suffering in others, would not be compelled to share our moral beliefs, as moral beliefs are dependent on affect.[4] Nichols writes, "Furthermore, if the affect-based account is right, their disagreement with us would not be due to any defects in reason or to the lack of ideal circumstances. Rather, their disagreement with us would flow from basic, arational differences in emotional response" (186). We might think that harming a person who poses no threat is wrong, but the aliens would disagree with us simply because they feel no distress upon causing innocent humans to suffer. Because the aliens in question do not have the same affective repertoire as us, and because of the crucial role that affect plays in determining our moral beliefs, there is nothing in virtue of which we can claim that the aliens must share our moral beliefs or affective repertoire.

Regardless of the strengths of Nichols' argument against moral objectivity generally, it provides us with good reason to think that sentimentalism is at least not germane to moral objectivity as he defines it. If moral norms are infused with feeling because of our natural responses to suffering in others, which causes us to believe that moral norms are not contingent on authority or individual preferences and desires, then unless we can

[4] Nichols lists psychopaths as evidence of people who, on account of their emotional deficit, simply do not share our moral judgments, and therefore provide a real-world example of the aliens he describes in this argument.

prove that people should have a certain affective repertoire, we cannot claim that moral objectivity is true.

More to the point, Nichols' attitude towards what he calls the 'persistence of moral judgment' is both telling and problematic. Based on experiments on undergraduate university students who reject moral objectivism, Nichols argues that people who reject moral objectivity still 'pass' the moral/conventional task. That is, they still identify norms against harming others as being non-authority contingent and as more serious than norms against, say, disrupting the professor during class, despite the fact that they are non-objectivists (195-196). Non-objectivists do not seem worried that their moral judgments are not justified by an objective morality. Indeed, Nichols writes as though they need not be concerned with justification at all:

> It seems presumptuous to say that giving up objectivity means that I should not judge harmful violations to be more serious, that I should not judge the wrongfulness of hitting another as independent of the teacher's authority, or that I should not think that the actions are wrong because they are harmful. As philosophical sentimentalists have long maintained, you do not have to believe that an action is objectively wrong to have a deep and abiding opposition to such actions. (196)

Nichols is probably right to say that we do not need to believe that an action is objectively wrong in order to be justified in our feeling strongly about it. However, perhaps there should be something in virtue of which we can say that we believe that something really is wrong in order to justify, to ourselves, the 'deep and abiding opposition' we feel towards it. Nichols remains silent on just what this something might

be, largely because he seems to force himself into a false dilemma: because we cannot secure objectivity, there is nothing left to be said by way of justifying our affective responses; it is objective justification or no justification at all. However, it seems intuitive that we cannot just take our individual emotional responses as given, and go on to argue that their phenomenological presence and psychological tie to motivation gives us all the reason we need to act or believe on the basis of them – our affective responses should not be taken to be self-justifying, as Nichols seems to imply. For, it seems that we can always ask whether it is good, right, beneficial, or even moral that we pass judgment and act on the basis of a specific affective response, and we even judge that we should or should not have certain responses in certain circumstances. We can reflect on our affective responses and in doing so we can criticize them as reasons for making normative moral judgments.

Perhaps some examples will help illustrate the point. In a study on the relationship between diet and perceived morality, researchers found that there is a measurable difference in how moral a person is perceived by others based on diet (Stien and Nemeroff 487). In this study, participants were given a description of a fake person that described some general characteristics about him or her, such as gender, height, weight, favoured physical activities, and eating habits. The findings indicate that participants generally rated people who eat fruit, salad and chicken as more moral than people who eat steaks, cheeseburgers, and fries.

In a study on the relationship between disgust and moral judgments, participants were hypnotized to feel disgust upon reading arbitrary words, such as 'take' and 'often.' They were then presented with a number of vignettes, some of which included the words 'take' and 'often.' These vignettes described scenarios where people ate their pet dogs, stole

library books, etc. There were also morally neutral vignettes; for instance, one described a student's council member, Dan, who tries to foster good discussions between professors and students about academic issues. As predicted, hypnotized participants rated the transgressions as more morally wrong when they felt disgust upon reading either 'take' or 'often' in the story. Interestingly, the researchers found that, even after they were aware of their hypnotized state, "some participants continued to follow their gut feelings and condemned Dan in the student council story, even though his only crime was trying to foster good discussions" (Wheatley and Haidt 783).

In both these studies, participants allowed their negative affective responses to things like unhealthy food and random words to influence their moral judgments. Clearly, should they reflect on the extraneous factors that affected their respective decisions, they would likely realize that their moral judgments were not an accurate reflection of their considered moral beliefs. If they did not, surely we want to be able to claim that the things some of the participants think are relevant to rendering moral judgments are inappropriate to that task.[5] A hard working student's council member such as Dan does not deserve to be morally condemned for such an arbitrary reason as the disgust one feels because of hypnosis!

We can reflect on the emotions we feel, and can offer criticisms and justifications of them. There are many subjectivist answers we can give by way of justifying or criticizing our affective responses that do not assume or push us toward an objective morality. Indeed, it is short-sighted to think that we should embrace the dictates of the emotional experiences we have only because of their individual phenomenological presence and

[5] Karen Jones makes this point albeit in a slightly different context, and with different aims. See her “Meta-Ethics and Emotions Research: a Reply to Prinz.”

motivational efficaciousness, especially if doing so encourages us to ignore our better judgment.

Self-Justification, External Justification, and Internal Justification

As we have seen, one problem with Nichols' reasoning about the 'persistence of moral judgment' lies in the way he approaches the justification of affective response. He seems to assume that because our affective responses cannot be given an objective justification, they must be self-justifying if we are to take them seriously. Recall that he thinks that there is no way to find out whether we have the right set of emotions or not (185), largely because "our emotional responses themselves have no *externally privileged* status" (184; my emphasis). Now, what does Nichols mean by 'externally privileged'? Since the job of the objectivist is supposed to be to show us that we all should have a specific affective repertoire, it is clear that, on pain of circularity, the objectivist cannot appeal to the dictates of his favoured affective repertoire when arguing that others should share that very same repertoire: "The objectivist cannot simply help himself to a moral intuition that rational creatures should have these emotions, because the Humean point is that our moral intuitions depend on the emotions we happen to have" (188). Thus, giving certain of our affective responses an externally privileged status would involve appealing to something *other* than our affective responses or the judgments we derive from them, for example, our shared rational nature or God's commands. Because Nichols cannot see an answer to this problem for moral objectivity forthcoming, he simply asserts that moral judgment persists, and that we do not need to be objectivists to feel comfortable in continuing to make such judgments. Therefore, our affective responses are, I take Nichols to imply, self-justifying. But as we have seen, this conclusion cannot be right. Is there

another way we can justify our affective responses other than by saying that they are self-justifying?

Nichols is not the only sentimentalist to take on the self-justification position. Perhaps a look at the work of Simon Blackburn, who appears to share much with Nichols' position on this matter, will help us make clear just what is involved in the self-justification position. In his article "How to be an Ethical Anti-Realist," Blackburn writes:

> Does the lover escape his passion by thinking, "Oh it's only my passion, forget it"? When the world affords occasion for grief, does it brighten when we realize that it is we who grieve? (The worst think to think is that if we are "rational," it should, as if rationality had anything to tell us about it.) [. . .] The news comes in and the emotion comes out; nothing in human life could be or feel more categorical. (175)

Blackburn's point is simple. The categorical quality of affective responses gives us all we need to believe or act on their respective behalves. Talk of rational emotions is superfluous. In the heat of the moment, thoughts about whether one's grief, love, or joy is rational are each just 'one thought too many;' trying to decide whether or not a felt emotion is rational takes away from the proper functioning of that affective response and is detrimental to it.[6] Thus, we should accept the categorical nature of our affective responses, and take them to be self-justifying. Or so I take him to imply.

Indeed, there is an important sense in which the seeds of normativity are present in affective responses *simpliciter*, and Blackburn captures this sense well. If something is

[6] Bernard Williams coins the phrase 'one thought too many' in his paper "Persons, Character, and Morality."

normative, for instance, a moral norm, it has a strong 'oughtness' to it; it compels us to do what it says. Affective responses seem to share this quality; they often compel us to do or believe. There is a certain sense in which affective responses are categorical, as Blackburn suggests. The sheer phenomenological force and, more importantly, the motivational efficacy of affective responses go far in making us feel normatively compelled to do or believe as they suggest.[7] While we still cannot say that affective responses are not *enough* on their own to provide justification, it seems right to say that by their very nature they do much of the normative work by themselves.

Notice, though, that Blackburn's examples of affective response, love and grief, are perfect for proving his point about the categorical nature of the emotions. Love and grief smack us in the face with demands; it is difficult to escape for their respective grips, and following the dictates of those responses is something that we would not usually question a person for doing. However, there are other affective responses that, while they share much the same categorical quality as love and grief, should probably not move us to believe or act so easily. Consider envy, anger, and hate; the presence of these responses may seem very categorical, but does that mean that they are self-justifying? If Ted feels hate towards someone and acknowledges that he feels hate towards that person for no appreciable reason, is Ted still justified in hating him? If affective responses are self-justifying, then the answer must be yes. But something seems horribly wrong with this reply; surely Ted is not justified in saying that someone is hateful only because at one point in the past he felt hate towards him, apparently for some reason that Ted cannot

[7] It is not unusual for philosophers to try to get at normativity through motivation. In fact, Stephen Darwall has proposed that placing normativity within the natural order requires understanding normativity through motivation: "For the philosophical naturalist, concerned to place normativity within the natural order, there is nothing plausible for normative force to be other than motivational force, perhaps when the agent's deliberative thinking is maximally improved by natural knowledge" (Internalism 168).

articulate. Hopefully, most people would criticize Ted for hating someone without reason, or even for feeling hate in the first place. But what could the basis for that criticism be?

We could attempt to develop an external standard whereby we can justify and criticize our affective responses. However, recall that I have argued that the phenomenological force and motivational efficacy of affective responses go quite far by themselves in establishing normativity. Now, if we are trying to develop a standard external to our affective responses *for* normatively evaluating affective responses, we are developing a new normative standard that must be wholly different from the *prima facie* normativity that seems to be inherent in affective responses.[8] If the external normative standard is not wholly new, then it would provide circular justification, which would defeat one of the purposes of developing a standard external to our affective responses in the first place.

Attempting to develop an external normative standard would seem to be an unnecessary endeavour for the sentimentalist; if our affective responses already provide us with fodder that might help us establish normativity, we should at least see if we can develop an acceptable normative standard with the normative pieces that our individual affective responses have already provided before attempting to develop a whole new standard. In addition, developing a whole new normative standard presents new problems in its own right. For example, it is unclear how such a standard could retain the motivational efficaciousness associated with normativity without being somehow tied to affect or desire, in which case it would not be a new standard at all.

[8] Christine Korsgaard makes this point in her "The Sources of Normativity:" "Morality must be endorsed or rejected from a point of view which itself makes claims on us and so which is itself at least potentially normative" (54).

This reasoning leads to an interesting objection against the search for an objective morality within a sentimentalist framework. Because an objective justification of our affective responses must be an external justification (see Nichols argument against objectivity, described in the preceding section), and because external justifications might be unnecessary, attempting to formulate an objective sentimentalism might be unnecessary. Of course, it is only unnecessary if one does not think that objectivity must be a defining characteristic of an acceptable moral theory. But if one does not think this way, to attempt straight away to develop an objective sentimentalism might lead us into error, because doing so could cause us to ignore other options that may in the end turn out to be more plausible.

Indeed, merely offering justifications for or criticisms of our affective responses does not necessarily push us toward objectivity. One can justify one's affective responses to another person without thinking that the other person must *necessarily* be compelled to accept that justification because of some irresistible considerations external to the affective response itself. The justifications we give are just as much for our own benefit as they are for the benefit of our fellow conversants. Nichols' error in approaching the problem of justification is that he seems to assume that this must be true: he assumes that any kind of justification of affective responses other than self-justification must be external and in the interests of securing moral objectivity. As a result, he neglects to investigate other forms of justification.

Perhaps, then, we should look to see if certain of our affective responses might be given some kind of *internally* privileged status, in the sense of internal to an individual's set of affective responses, the moral values and judgments derived from those responses,

and, when they are relevant, any other motivating practical claims that apply to that individual. While obtaining this status for some our affective responses might not lead us to having the 'right' set of them, we might be led to having the 'best' set of affective responses, in the sense of 'best relative to the individual (or group of sufficiently similar individuals) under consideration.' To anticipate, the prime methodological candidate for attempting an 'internal justification' for our affective responses will be a suitably precise formulation of the notion of coherence. I will return to this topic in chapter three.

Let's go back to the hating example for a moment, to see whether an internal standard of criticism would allow us to criticize Ted for his hatred. In virtue of what could a person criticize Ted for hating someone, just because at one point Ted felt hatred for that person? Well, let's assume that at some point in the past Ted has said that hating other people accomplishes nothing valuable. He knows that nothing he *feels* is valuable can come from hating another person. Now we have a clear reason for criticizing Ted: he is inconsistent. He has said that he believes that nothing that he feels is valuable can come from hating a person, but he also insists on hating one person in particular. In effect, Ted has conflicting affective responses toward his own hatred - he feels compelled to keep hating someone, yet he also feels that he should not. Unless Ted can, in good conscience, find a way to resolve this conflict, he should be compelled, on pain of inconsistency, to try at least not to hate the person in question. In short, there seems to be a method whereby we can justify and criticize our affective responses without pursuing an externalist course, while still avoiding the pitfalls of self-justification.

To summarize, the self-justification position, while it sheds important light on the inherent normativity of affective responses, fails to deal with the fact that having an

affective response is not sufficient for justification. However, this does not entail that the theorist concerned with the justification of affective responses must attempt to find an external justification for them; we should first try justifying them internally, in the sense of internal to an individual's set of considered affective responses, moral values, and judgments.

Conclusion

In the foregoing, I have tried to argue for three general points. First, even though the simple sentimentalist arguments in favour of a more inclusive definition of moral judgment are for the most part correct, they ignore an important subspecies of moral judgment, namely, normative moral judgment, which deserves consideration in its own right, both descriptively and normatively. Second, while Nichols argument against moral objectivity is very promising, his view about 'the persistence of moral judgment' is unsavoury insofar as it seems to imply that affective responses are self-justifying. Third, after fleshing out Nichols' self-justification theory, and showing how it is similar to Blackburn's views on the same topic, I argued that the while affective responses taken by themselves *do* have substantial normative authority, the self-justification account is quite limited insofar as it cannot account for cases where self-justification is not enough, both descriptively and normatively. However, rejecting the self-justification of affective responses does not drive us to attempt to try an external justification; rather, I argued, there is a third way: internal justification.

Empirically informed, descriptive moral theories such as Nichols' are no doubt important to the study of ethics. They provide interesting insights and in many important ways constrain our theorizing. However, the normative questions cannot be ignored and

certainly are an important part of any ethical theory. An acceptable sentimentalist ethical theory should provide not only understanding about the subject matter of ethics but also some guidance about how to separate reason-giving affective responses from non-reason giving responses. It would have to state the conditions upon which an affective response can be justified as a reason for belief and action for an individual. In other words, an acceptable sentimentalist theory must at least make an attempt to be normatively adequate.

Sentimentalist Normative Moral Judgments: Replies to Objections

In the first chapter, I argued that empirically informed sentimentalist moral theories tend to focus on the descriptive aspects of moral theory, and that this emphasis causes their moral theories to be problematic when applied to normative questions. I also argued that an internal method of justifying affective responses as normative reasons is the best form of justification for sentimentalists, while anticipating that the best method for justifying affective responses is a suitably precise notion of coherence. In this chapter, I will take a slightly different approach. In order to better understand how affective responses can function as justifying reasons for normative moral judgments, I will attempt to respond to two challenges posed for sentimentalism by Stephen Darwall, Alan Gibbard, and Peter Railton.

In their review paper "Toward Fin de Siecle Ethics," Darwall, Gibbard, and Railton point out two distinct but related problems for accounts of moral judgment that implicate the emotions. The first goes like this: "Emotivists hold that a moral judgment consists in a feeling – or better, in a disposition to have certain feelings. It seems, though, that a person can judge something wrong even if he has lost all disposition to feelings about it" (149). This argument presents a challenge: if moral judgment is supposed to consist in a feeling or a disposition to have certain feelings, how can people make moral judgments dispassionately? If a person does not feel the force of an affective response when making a moral judgment, or does not appear to be at all disposed to feelings about the specific content of their moral judgments, then it is not apparent that such judgments consist in the feelings associated with affective response, or even with dispositions to have certain feelings.

The second problem is slightly more involved:

> What then is this feeling of moral disapproval? Among theorists of emotion, cognitivists dominate. Emotional "cognitivism" is different from meta-ethical cognitivism: an emotional cognitivist thinks that having a certain emotion, such as anger, involves making some special kind of cognitive judgment. Now in the case of moral disapproval, the only plausible candidate is a cognitive judgment that the thing in question is morally wrong. If so, we need to understand judgments of wrongness before we can understand moral disapproval. We cannot explain the judgment that something is wrong as an attitude of moral disapproval. (149)

Even if the emotions are implicated in moral judgment, the authors imply, they involve cognitive judgments. Hence, it isn't the feeling that provides an explanation, but rather the cognitive judgment implicit in affective response, and therefore we should focus on the cognitive judgment rather than the 'feel' of affective responses.

Both problems are important objections to an account of normative moral judgments that are justified by affective response, and I will respond to each in turn. Fortunately, a review of some more recent work in neurology and psychology should help decrease their potency. In addition to helping meet the aforementioned objections to the account of moral judgment on offer, this quick review will also help us better understand two very important things: first, in relation to the first objection, how people can reflect on and reason about their emotions, without being biased or overrun by an emotional experience, and second, in relation to the second objection, how affective responses can be thought of

as functioning as the normative component of moral judgments.

Objection One: Dispassionate Moral Judgment

The first objection poses a challenge to explain how we can make moral judgments dispassionately, or in the absence of any disposition to have affective responses, while still explaining those judgments in terms of affective response. My approach to this problem will be twofold. First, I will attempt to show that people can be mistaken about whether or not they are in an emotional state or are disposed to have certain affective responses. Put differently, I will argue that a person can mistakenly believe that they have lost any disposition to feeling toward some object. The primary motivation for this argument is that there is difference between how psychologists and neuroscientists talk about affective response, and how most people conceptualize emotional feeling or being in an emotional state. Once this has been clarified, I will turn to the problem of people who seem to make moral judgments but *actually* have lost all disposition to feelings toward the objects of their moral judgments.

As it turns out, we can help explain why people can be mistaken about their own emotional states and dispositions by reference to a neurologically-based theory of the emotions, initially developed by Antonio Damasio, called the 'Somatic Marker Hypothesis.' Put simply, the determining idea behind this theory is that "decision making is a process which is influenced by marker signals that arise in bio-regulatory processes, including those that express themselves in emotions and feelings" (Bechara 25). Emotions are defined in terms of somatic state, because "(a) emotion induces changes in the physiological state of the body, and (b) the results of emotion are represented primarily in the brain, in the form of transient changes in the pattern of activity in the

somatosensory structures" (Bechara 6). These 'marker signals,' which are tied to representations of objects and events in our environment, express themselves as emotional states that provide us with evaluative information about how to interact with our environment. Therefore, the hypothesis suggests, the emotions perform a crucial role in decision making.

The aspect of this theory relevant to the matter at hand is its postulation of two types of affective response. One type of response is called the 'body loop.' In this chain of events, the emotion is realized both in the body proper and in the relevant subcortical and cortical processing structures in the brain (Bechara 6). The second type of response is called the 'as if body loop.' Interestingly, this chain of events is supposed to occur after we have already experienced and expressed an affective response in the body loop, or, in other words, after we have 'learnt' an affective response. After this has occurred, affective responses can "bypass the body altogether, activate the insular and somatosensory cortices directly, and create a fainter image of an emotional body state than if the emotion were actually expressed in the body" (Bechara 7).

Now, if a person is making a moral judgment based on an affective response that has been enacted in the 'as if' body loop, the *feel* of the affective experience that motivated that judgment would be weaker than that associated with judgments made based on affective responses that involve the body loop (Damasio 155-158). Importantly, it would still contain some representation of a bodily state, even if the body does not represent an actual bodily state. It might even be the case that certain of our affective responses are so well 'learnt' that the feelings associated with them could be quite faint, though there is not, to my knowledge, specific data on this point. On this theory an affective response

can still motivate a moral judgment, but the level of feeling associated with that response could be faint enough that it differs significantly from the way most people would conceptualize an emotional experience. Having this type of affective response might not cause people to think that they are in the grips of what they would typically conceive of as an emotional experience.

Another relevant distinction involves the way in which affective responses are triggered. 'Primary inducers' are things that directly cause a response, for instance, a growling bear or first reading the letter that tells you whether or not you have won an important scholarship. 'Secondary inducers' are internally generated by recalling an emotional event, or concocting a hypothetical emotional situation, such as visualizing an angry bear (Bechara 12). Both types of inducers can bring about a response either in the body loop or the 'as if' body loop. For example, a secondary inducer, such as recalling the death of a loved one, can certainly cause a strong response enacted in the body loop. On the other hand, a primary inducer, such as looking at the ground while rock-climbing, could be so well-learnt for some people that it causes a weaker response that is only enacted in the as if body loop.

Despite the fact that secondary inducers can bring about a strong affective response, it seems reasonable to postulate that secondary inducers that a) have been learnt by repeated experience and b) are not associated with, or have been disassociated from, any strong and personal emotional experiences would likely result in a weaker emotional response. Now, in many instances moral judgments are internally generated on the part of the person making a judgment, i.e. they arise 'in thought.' Because they are internally generated, these judgments would involve a secondary inducer. Also, making a moral

judgment often involves making some kind of general claim, such as 'murder is wrong,' which, in many cases, would not be associated with a strong, personal emotional experience.[9] So, if one is in the habit of making a certain moral judgment based on a secondary inducer that is not associated with a strong, personal emotional experience or memory, when one makes that judgment, it will likely be accompanied by a weak affective response. Consider the way philosophers toss around judgments like 'murder is wrong' in seminar and colloquium discussions. In those contexts, standard examples of moral judgments function more as variables for moral judgments than as moral judgments themselves, even on the part of those who introduce such examples. The upshot is that it is no surprise that the level of feeling that people tend to associate with an emotional experience is not implicated in such judgments, although some feeling, however faint, still is.

Remember, the challenge is to show that we can still explain a person's moral judgment in terms of affective response, even when the person in question "has lost all disposition to feelings about it" (Darwall, Gibbard and Railton 17f). Thus far, I have shown that under certain conditions people can make moral judgments that only implicate weak affective responses, which entails that people can make moral judgments without being in what they might call an emotional state while making that judgment, thus confusing them about their actual dispositions. In a sense, I have only partially responded to the challenge, because I have not shown that affective response is still implicated when a person makes a moral judgment about something that he has lost *all* disposition to feeling about it.

[9] With cases like 'murder is wrong,' the affect involved that motivates the judgment might even be associated more with social pressures rather than with actual aversion to doing another person harm, i.e. people usually don't have to murder someone only to find out that it is wrong; we learn these types of things more by social reinforcement, it would seem.

Now, if a person has lost all disposition to feelings towards something, they must have at some point been disposed to feelings towards that thing. Recall the psychological distinction make in the first chapter between conventional and moral rules. An important difference between the two kinds of rules is that moral rules always involve some kind of victim, and thus involve affective responses to harming others, whereas conventional rules do not (although, of course, conventional norms can involve different types of affective responses). I do not want to make the argument that all moral judgments necessarily involve some kind of victim, but it is at least reasonable to assume that a very important subset does involve victims and harm norms.

With this background in mind, Shaun Nichols makes a good point: "It is likely that core moral judgment can persevere for at least some time after the emotions are eradicated. But it is also likely that over time, the tendency to treat harm norms as distinctive would wane" (99). So, if a person lost all disposition to feelings toward, say, striking another person, but still made judgments like 'striking another person is wrong,' that judgment would be more like a statement of a conventional norm than a moral norm. Now, I do not want to make an argument about whether such judgments 'really' are moral judgments. But assuming that they are, such judgments would be importantly different from the moral judgments about harming others that normal people make, especially insofar as they would probably present themselves to their speakers as being contingent on authority, which implies that they would not be as motivationally effective. So, while someone might be able to make a moral judgment while having lost all disposition to feelings about the object of that moral judgment, the way that judgment would fit into a person's moral reasoning, and the extent to which they would take that judgment as a

constraint on their behaviour, would be severely compromised.[10]

Thinking about Affective Response

Even if we are able to make moral judgments dispassionately, how does that reflect on the sentimentalist hypothesis in general? Should sentimentalists not want it to be the case that people cannot make 'real' normative judgments without being in the grips of some strong emotional experience? To the contrary: our ability to ‘distance’ ourselves from our affective responses, and our ability to evaluate whether affective responses actually provide us with normative reasons from that distance, are crucial when attempting to establish whether *prima facie* reasons for action and belief, in this case affective responses, are in fact normative reasons for us. The reasoning behind this claim is simple. If every time we tried to think about whether some affective response to some stimuli provides us with a reason for making a normative judgment, we were overrun by a strong emotional experience, our decisions would always be biased because we would be in the grips of a strong emotional experience. For example, if every time you tried to decide whether anger was appropriately directed at some person or state of affairs, you were overcome by anger, it would be difficult to judge whether that anger provides you with a good reason to pass moral judgment, because your natural inclination would be to say that your anger is in fact justified.

The existence of a body loop and an 'as if' body loop helps answer this concern. When an affective response is instantiated by the 'as if' body loop, the feeling that accompanies it is much weaker than that that goes along with responses instantiated in the body loop, which suggests that once we have learnt an affective response, we are able to induce that

[10] While not an example of people who have lost all disposition to feel, psychopaths provide an example of people who never have had a disposition to feel (Blair et al. 2005). In line with this argument, psychopaths have significant problems when distinguishing between moral and conventional norms (Blair 1995).

response without putting ourselves in the grips of it. Put another way, this theory indicates that we can think about our affective responses without being overcome by them.

In their paper "Cognitive Regulation of Emotion: Application to Clinical Disorders," Pierre Philippot et al. propose a theory of cognitive regulation of emotion that also helps to answer this concern by giving a more detailed description of how people might be able to think about their affective responses and the things that trigger them. This theory postulates that there are two ways of experiencing emotion: actually experiencing an emotion, and being aware of the fact that one is experiencing or has experienced emotion. Philippot et al. label the first type of experience 'noetic', and the second type 'autonoetic' (78). The theory also postulates that there are two ways in which emotions can be represented in the brain: in the form of propositions, and in the form of schemata. When represented as propositions, the emotions can be understood as declarative knowledge about the emotions (76). When represented as schemata, the emotions are best understood as "an abstract and implicit representation which integrates sensory, perceptual, and semantic information typical of a given category of emotional experiences, on the one hand, and their relation to the activation of specific body response systems, on the other" (75).

Since we are talking about how a person can think about her affective responses and their causes, the type of emotional experience that we are most interested in is the autonoetic variety, as it involves us being aware of the fact that we are having or have had some emotional experience. On the topic of autonoetic experiences, the authors note that if each time one recalled some past emotional experience, one actually re-lived that

experience by activating the relevant emotional schema, "the whole cognitive system could be disrupted by the emotional arousal and would start to dysfunction" (79). Therefore, they argue, "autonoetic consciousness processes should have the capacity to regulate the activation of the schema to obtain only the schematic information needed, without risking potential disturbances for high level cognitive processes" (79). Therefore, on this theory, when a person reasons about his affective responses and their contents, he should be able, to some degree, to control the level to which the schema associated with that affective response is activated, and hence to be able to think about their affective responses, if he so chooses, relatively dispassionately.

Importantly, Philippot et al. propose that one of the major functions of autonoetic awareness of affective experiences is to inhibit the activation of emotional schemata (and, conversely, allow its activation to continue). Thus, if a person is aware that she is having an affective response, if she judges that that response is not justified, she is able to control the extent to which she *feels* that response.

The upshot of all this is that there is good reason to think that people can profitably think about their affective responses. Of course, the somatic marker hypothesis and Philippot et al.'s theory do not prove this decisively, but rather only give us good reasons for believing it is true. As far as sentimentalists are concerned, this is an important conclusion, as it states that people are not 'held hostage' by whatever affective responses that they happen to have. If we did not have a requisite amount of control over our affective selves, attempting to decide which affective responses are justified and which are not would not be a productive exercise, as we would not be able to implement any conclusions we might reach. Fortunately, this does not seem to be the case.

Before moving on to the next objection by Darwall et al., a potential inconsistency needs clarification. In chapter one, I argued that the phenomenological presence and motivational efficacy nomologically associated with affective responses gives them their *prima facie* normativity. This seems to conflict with what I have argued in the preceding, wherein I have stated that one does not have to have an emotional experience, traditionally conceived, as a consequence of having an affective response upon which one bases a moral judgment. I also argued that the 'feel' or phenomenological presence implicit in an affective response can be weaker than what people would normally associate with having an emotional experience, and therefore, in the case of moral judgments, people might mistakenly believe that emotions or affective response do not ground their moral judgments. In the first chapter, I talked about the importance of phenomenological presence and motivational efficacy, and in the preceding it might seem that I have argued for the opposite.

First, we can meet this inconsistency by claiming that in 'no-pressure' situations (e.g., in seminar), that moral judgments only implicate a disposition to have a phenomenologically strong affective response in certain situations. So, when I say 'murder is wrong,' that statement is still grounded in affective response, because, I am claiming, it relies on the fact that if I were in a situation where I was confronted by a murder, or was contemplating doing some such immoral act, I would have a strong, aversive emotional reaction. In addition, it is not as though no affective response at all is implicated in judgments made in 'no-pressure' situations; such judgments still would implicate a weak affective response.

Second, there is a difference between merely uttering a moral judgment in a ‘no

pressure' situation and questioning whether an affective response can function as a justifying reason for a moral judgment. For, in the latter case, what one is grappling with just is one's affective responses, which of them one would feel in hypothetical situations, or even those which one has felt in emotionally memorable situations. Thus, when considering whether some judgment is properly motivated by an affective response, the pertinent factors indeed are phenomenological presence and motivational efficacy.

Objection Two: Affective Response, Cognitive Judgments and Moral Judgments

The second challenge posed by Darwall, Gibbard and Railton involves the nature of the emotions themselves. If having an affective response involves making a special kind of cognitive judgment, they argue, we cannot explain moral judgments in terms of the affect (or, in their words, 'attitude') inherent in affective response, because that feeling would have been caused by the special kind of cognitive judgment in question. We cannot explain the (normative) judgment that something is wrong by referencing an affective response whose impetus is a cognitive judgment that something is wrong (18). In the following, I will argue that with regard to at least one important class of moral emotions, this approach is correct. Moreover, I will argue that this way of looking at the relationship between moral judgments and affective responses is an oversimplification.

In posing this objection, Darwall et al. refer to the dominant position of cognitivists among emotions theorists. In order to meet their objection, we need to pin down what being an emotions cognitivist in this context means. Perhaps they are talking about what Patricia Greenspan describes the evaluative, belief based approach to the emotions:

> As applied to accounts of the nature of emotions, the evaluative approach is often discussed under the heading of "cognitivism," which interprets

> emotions themselves as amounting to or containing cognitions, usually seen as mental states representing evaluative propositions. The most straightforward version of the cognitivist view is "judgmentalism," which understands the nature of emotions in terms of judgments: assertive propositional attitudes, assessable by ordinary evidential standards applied to beliefs. (Practical Reasoning 209)

Darwall et al. state quite clearly that understanding a judgment of wrongness is a prerequisite for understanding an attitude of moral disapproval: "Now in the case of moral disapproval, the only plausible candidate [to explain the attitude] is a cognitive judgment that the thing in question is morally wrong. If so, we need to understand judgments of wrongness before we can understand moral disapproval."

Hence, Darwall et al. seem to be claiming that feelings of moral disapproval do not explain/cause moral judgments, but rather than moral judgments explain/cause feelings of moral disapproval. My approach to this problem will not be to argue that the chain of causal explanation necessarily runs the other way. Rather, what I hope to show is that this view of the relationship between judgments of wrongness is at best an oversimplification. The best way to accomplish this is to consider the case of psychopaths, who are severely diminished in their capacity to generate affective responses to the fearfulness and sadness of others (Blair et al. 1997, 2005; Blair 1999). Psychopaths are also notoriously prone to antisocial and immoral behaviour, and the likely cause of this behaviour is indeed their reduced ability to generate affective responses to the distress of others (Blair et al. 2005, ch. 8). If we apply the cognitivist approach to emotions favoured by Darwall et al. to this case, the root cause of psychopaths' reduced ability to generate emotional states should

manifest itself in a reduced ability to make certain evaluative judgments, specifically, judgments of moral wrongness. This brief case study will show us two things: first, that judgments of wrongness and feelings of moral disapproval are not per say the problem with psychopaths, but rather that judgments about the distress of others, and affective reactions toward the distress of others, are more so the problem. Second, and more importantly, it will show that the 'appraisals' that are involved in eliciting what are properly called emotional reactions, do not present themselves as judgments in any traditional sense.

When tested against their ability to distinguish between transgressions and norms according to the moral/conventional divide, psychopaths fare quite poorly. They have trouble distinguishing between the two types of social transgressions. Furthermore, they are far less likely to give justifications for moral judgments that involve victim's welfare. For example, when asked why it is bad to hit another person, control subjects usually justify this judgment by reference to the victim's welfare, e.g. they cite the distress it causes the victim by saying something like 'it hurts him.' In contrast, psychopathic individuals are far less likely to give victim's welfare justifications; in fact, they are *more* likely than control subjects to give *normative* justifications than control subjects, e.g. by saying 'it's wrong to X' (Blair 1995). This indicates that the root problem with psychopaths' difficulty with the moral/conventional distinction is not necessarily their inability to generate judgments of wrongness. Rather, their reduced ability to consider and react to the distress of others, e.g. others' sadness and fearfulness, seems to be the pertinent causal factor behind psychopaths' poor performance in the moral/conventional task. Hence, the problem presents itself more so through their reduced ability to consider

and react to the distress of others, e.g. others' sadness and fearfulness, than it does through any general deficit in evaluative judgment or moral judgment, contrary to what emotions cognitivism predicts.

Perhaps, then, psychopaths have a reduced ability to make judgments about the distress of others, i.e. they are less able to judge that someone else is sad or fearful. However, the data on this point is unclear. For example, a series of studies have tested psychopaths' ability to categorize the emotional expressions of other people have produced inconsistent results. For instance, in one study, researchers found that psychopaths were able to categorize pictures of sad and fearful facial expressions equally well as non-psychopaths (Kosson et al.). In another study on psychopaths' ability to attribute complex mental states (e.g., doubtful, upset, insisting, despondent, etc.) to other people, researchers found no difference between psychopaths and control subjects (Richell et al.). However, in study on children with psychopathic tendencies, Blair et al. found that psychopathic children's' ability to recognize both fearful and sad facial expressions is impaired compared to comparison children (2001). Upon running the very same facial affect recognition experiment on psychopathic adults, however, Blair et al. found that psychopaths were less able than comparison individuals to recognize fearful, but not sad, facial expressions (2004). Despite these contradictory data, some authors have pieced together an interesting explanation

Firstly, why do psychopathic children perform more poorly on facial recognition tasks than psychopathic adults? One prominent explanation for this discrepancy points to a possible root cause for psychopathy. In non-psychopathic individuals, the processing and recognition of sad and fearful expressions involves recruiting certain sub-cortical

systems in the brain, especially the amygdala. Psychopathic individuals have reduced amygdala volume, and show reduced amygdala activation when processing sad and fearful expressions (Blair et al. 2005, 116). Thus, while psychopathic children are likely initially impaired in their ability to recognize sad and fearful expressions because of their amygdala deficiency, it has been suggested that psychopathic adults have been able to learn to compensate for their deficiency over time by recruiting other areas of the brain to help process this data, namely, cortical areas.

Now, when the brain relies on areas of the cerebral cortex, or cortical areas, to compute information, those processes tend to be slow, effortful, and time consuming. In addition, the cortex implements high-level cognitive processes such as language, reasoning, thinking, and consciousness. In contrast, when the brain relies on sub-cortical systems to compute information, these processes are quick and efficient, and they tend to occur automatically. Automatic processes are typically defined by the following criteria: "An automatic process is one that (a) can operate efficiently, (b) occurs without awareness, and (c) is difficult to control" (Blair, I. et al. 764).

Another look at the inconsistent data shows an important pattern. When adult psychopathic individuals are given an easy facial identification task (e.g. Kosson et al.), or are given ample time to respond to a facial identification task (Richell et al.), they perform equally well as, if not better than, comparison individuals. In the first case, because the task was not overly demanding, the subject’s cognitive resources were able to compensate for their amygdala deficiency. In the second case, because the subject was allowed to complete the experiment at his own pace, his cognitive resources again would not have been strained, and he would have been able to compensate. On the other hand,

in Blair et al.'s 2004 study, when the subject's cognitive resources *were* strained by the difficulty of that task and precise time constraints, the psychopathic subjects were slightly slower to identify others' fearful facial expressions as fearful, but, in contrast to psychopathic children, they showed no significant difficulty with sad faces. In fact, because the psychopaths were slower in identifying fearful facial expressions, and slow processing is a hallmark of cortical processing, this supports the idea that psychopaths are utilizing higher-level cognitive processes in order to discriminate between different emotional facial expressions.

So, it seems that while adult psychopaths do show some impairment when recognizing fearful facial expressions, they are able to correct for the more prominent deficiencies found in psychopathic children by learning to recruit cortical processing structures to make those distinctions. In other words, psychopaths have to rely on areas of the brain associated with reasoning, language, and high-level thinking to complete this task, which are just the areas of the brain that would be involved in making evaluative *propositions*, and which, presumably, emotions cognitivists would propose play an important role in eliciting the feelings associated with affective responses. However, adult psychopaths still present with reduced affective responses to the fearfulness and sadness of others, which presents a problem for Darwall et al. insofar as they rely on a heavily cognitivist theory of the emotions. For, psychopaths present with no general impairment in their ability to make high-level evaluative propositions *despite* amygdala dysfunction. Although they can still understand *that* a person's face expresses sadness, or *that* hitting another person is bad because it is wrong, etc., they remain impeded in their ability to generate genuine *felt* affective responses.

Indeed, the problem with psychopaths does not lie in their high-level cognitive systems, but in the low-level, automatic, sub-cortical cognitive systems involved in eliciting what are properly called emotional, or affective, responses (Zhu and Thagard 29).[11] Now, in some way the processes that these sub-cortical structures implement are involved in making crude appraisals of stimuli that occur causally prior to felt responses (Zhu and Thagard 27-29). In other words, one of the sub-cortical process that elicits felt emotional responses 'decides' whether some stimulus is 'good' or 'bad' for the organism. Because we have seen that high-level evaluative propositions do not help explain felt affective response, perhaps instead we can try to understand these automatic 'appraisals' as the evaluative judgments in question, as they are involved in an explanation of affective response.

Now, automatic appraisals of emotional information understood in this way neither present themselves propositionally nor do they appear to be executed in a propositional fashion, which means that they do not resemble propositional, evaluative judgments. There are two main reasons for making these claims. The first points out that the way these appraisals seem to present their 'results' for conscious processing is primarily in the form of felt affective response. The second involves the processes that instantiate these appraisals. The relevant sub-cortical processes do not seem to reason that 'is a threat' or 'is morally blameworthy' are proper predicates for some stimulus. Rather, they are more like systems that simply *respond* in a certain ways to certain stimuli, thereby resulting in a felt

[11] Even though I refer to these sub-cortical processes as 'cognitive,' it is important to realize that they are still automatic and intuitive. When we are talking about a psychological process, saying that it is 'cognitive' is usually only intended to mean that it involves information processing, whether it is affect-based or non-affect-based (see Blair J. 67). We cannot take the word 'cognitive' here as being opposed to 'emotional,' as one might be naturally inclined to think, because both types of process involve information processing.

affective response, which in turn can cause an evaluative judgment to be made.[12] In other words, these appraisals do not directly render propositional, evaluative judgments in any traditional sense.

To summarize, the evidence garnered from an examination of several studies on psychopaths suggests that high-level cognitive processes are not required for the generation of what are properly thought of as emotional responses, as emotions cognitivism, and especially emotions judgmentalism, would predict. Rather, low-level, sub-cortical cognitive processes are implicated in the generation of genuine affective responses. Psychopaths present with a significant deficit in this system (specifically, in the amygdala) which causes them to be far less emotionally responsive to the sadness and fearfulness of others when compared to control subjects. These sub-cortical processes, at one point or another, involve making appraisals of emotional information, and these processes do elicit felt affective response. However, these appraisals do not resemble evaluative judgments in any traditional sense. Clearly, though, such appraisals are very relevant to explaining affective responses and the judgments that people base on them, as they are causally prior to felt affective reactions (Zhu and Thagard 27).

Given all this, it is a gross oversimplification to say that we need to understand judgments of wrongness before we can understand moral disapproval. First, with regard to psychopaths, it seems that rather than talking about their emotional deficit in terms of 'feelings of moral disapproval,' we should be talking in terms of emotional responsiveness to the emotional states and distress of others.[13] Second, if we concentrate on the affective

[12] It is of course perfectly reasonable to talk about these systems 'categorizing' or 'appraising' something as a threat, for example. But this is more a manner of speaking than an actual description of what these systems are doing, I would think.

[13] Of course, there are other morally relevant emotions, e.g., guilt, anger, etc.

responses that are caused by the distress of others, it does not seem that any kind of evaluative judgment, traditionally conceived, is involved in eliciting those responses; instead, the closest thing to evaluative judgments that could be said to elicit such responses is some kind of automatic appraisal. Therefore, for at least one class of morally relevant affective responses (i.e. reactions to the distress of others), it is not apparent that evaluative propositions perform any significant role in eliciting those affective responses. Hence it is not apparent that they can explain the moral judgments that people make on the basis of those responses, as Darwall et al. claim in their objection.

At this point, there are two further problems that need addressing. First, there still remains the fact that automatic, sub-cortical appraisals are indeed implicated in the elicitation of affective response, and that in order to understand moral judgments based on those responses, we still have to understand automatic appraisals – concentrating on felt affect as a causal explanation of moral judgments is still not a worthwhile exercise. Secondly, I have not yet talked about the voluntary, evaluative judgments that people do in fact make in order to regulate their affective responses.

With regard to the first problem, because automatic appraisals occur unconsciously and involuntarily, whereas the affective response that these appraisals result in is felt consciously and is thus clearly more relevant to voluntarily passing judgment, it seems that the part of this process that is most relevant to the moral judgments we consciously and voluntarily make is the conscious feelings and motivational pushes and pulls associated with affective responses. Despite the fact that unconscious cognitive appraisals are plainly relevant to giving an objective *explanation* of moral judgments, such appraisals are hardly relevant from the perspective of the person trying to make

normative moral judgments, or so it would seem.

Now what about appraisals that are made voluntarily and propositionally, for determining future affective responses and behaviour? In other words, what about the judgments that we explicitly make regarding regulating our affective responses? Now, if a person voluntarily judges that some affective response is unjustified in relation to some object or state of affairs, we cannot just say that since this person has made a cognitive judgment that is supposed to determine future affective responses and behaviour, affective response does not play an explanatory role in the original judgment. Consider the following example. Sarah, who owns a coffee shop, has recently decided that she is conflicted about the positive affective response she feels when she makes a good profit, because she knows that the coffee beans she uses are grown by Latin American farmers who work as de facto slaves, thereby suffering greatly partly so that she can sell cheap coffee. In this case, because she bases her judgment on her aversion to the suffering of others, which she thinks just is the result of having negative affective responses to the suffering of others, affective response does indeed play an explanatory role in her judgment, for that is what motivates her to decide that she can no longer feel happy when she makes a good profit.

To summarize, if the cognitive appraisal that causes an affective response is made unconsciously and involuntarily, it does not resemble a judgment of wrongness traditionally conceived. Also, it seems that it is not unconscious appraisals that matter to people when they make normative moral judgments, but rather the feeling that these appraisals cause. On the other hand, if a person consciously makes a cognitive appraisal of some affective response that is intended to change his future affective responses, it is

quite possible that his appraisal was based on some other affective response.

Explaining the Normativity of Affective Response

In the foregoing section, I briefly argued that while the involuntary, unconscious appraisals that determine affective responses certainly are relevant to giving an explanation of how an affective response is implicated in an explanatory account of moral judgment, the feeling and motivational efficaciousness associated with affective response is what is relevant to the person actually making a normative moral judgment. Surely, though, if we know this, should we not take those involuntary appraisals more seriously when voluntarily rendering *normative* moral judgments, as they are, in some sense, the root cause of those judgments? The idea here is that when someone is making a moral judgment, instead of doing so on the basis of those aspects of his affective responses that surface in conscious experience, he should focus on the appraisals that actually cause the conscious aspects of affective response, and factor them into his decisions. I will argue that appraisals of this nature provide poor material for discussing and evaluating normative moral judgments.

There are a number of problems with this line of reasoning. For, if we are going to base our moral judgments on the automatic appraisals that cause affective response, we have to be able to, in some way, analyze those appraisals to see how they can help us in making normative moral judgments. We would have to know what basis there is for those appraisals, i.e., in what 'terms' these appraisals are made, or, 'why' they are made. This would be a difficult task. The operations of automatic processes are not consciously accessible, in other words, they are inaccessible from the first-person perspective (Haidt 6). Because they are not consciously accessible, we are only aware of the results of

automatic processes like these, not the processes themselves – we are aware of the affect and not the underlying appraisal. When appraisals are consciously made (i.e., when they are proper judgments meant to regulate one's future affective responses), their relevance to making moral judgments is relatively obvious. Consequently, in the interesting case, we cannot analyze these appraisals from the first person perspective.

The other option, then, is to try to analyze involuntary appraisals from the third person perspective. We could to try to understand them as systems or processes implemented in the brain. We could analyze them as neural systems or as systems picked for by natural selection. If we try to understand them as neural systems, the relevance of having this understanding to making normative moral judgments is only partial. Consider an example. Steve is wondering whether his feeling of moral outrage toward Guy is justified, and he is wondering whether knowing more about his unconscious appraisal process will help him to decide. The only advice neuroscience could give him would recommend that he does not commit an error by saying, for instance, that his moral outrage is justified by the fact that his automatic appraisals are rendered by automatic processes that make decisions according to the categorical imperative, as this would be an incorrect analysis of these appraisals (e.g. because they do not appear to make decisions in terms of the categorical imperative).

If we proceeded to tell Steve that when he feels outrage, this or that area of the brain is active, that the neurons involved release certain types of neuro-transmitters, that the process is automatic and intuitive, and so on, this information would not be relevant to whether his moral outrage is justified, as they would just be descriptions of how his feeling comes about. He is asking if he should continue to feel that way, and whether he

should pass judgment on the basis of that feeling. Neuroscience only tells us how we should *not* analyze our unconscious appraisals. Consequently, it shows us how we should not try to use them as justification, and it does not show us how they might be relevant to making normative moral judgments.

At this point we turn to evolutionary accounts that explain why we make the automatic appraisals that we do, to see whether they can provide us with a way to profitably analyze them. As an example of this type of account, neurobiologist Joseph LeDoux offers a reasonable description of the evolutionary advantage of automatic fear responses: "From the point of view of survival, it is better to respond to potentially dangerous events as if they were in fact the real thing than to fail to respond. The cost of treating a stick as a snake is less, in the long run, than the cost of treating a snake as a stick" (LeDoux 165.; cited in Zhu and Thagard 29). The appraisals that cause morally relevant affective responses are amendable to the same type of account. In effect, we make the (morally relevant) appraisals that we do because making them increases our ability to survive and reproduce. In this way, we can analyze our unconscious appraisals as products of natural selection.

With the foregoing in mind, if we try to understand the kind of cognitive appraisal that produces conscious emotional experiences, which is supposed to help us better understand how we should justify our affective responses, we can use an evolutionary account. If we apply this kind of account to Steve's case, we find that it is overly general, and that it likely conflicts with his own normative aims. For, we could only tell him that an appraisal process that was selected for elicited his feeling of moral outrage, which tells him that for the most part his episodes of moral outrage tend to be fitness enhancing. It

does not tell him much about his specific situation, i.e. it tells him nothing about whether Guy has committed a moral transgression. Also, Steve likely rejects the claim that evolutionary accounts of his automatic appraisals grant them any kind of normative authority – he might think that there are more important things than evolutionary fitness. Christine Korsgaard makes this point well: "as a moral agent, you might decide that moral claims, if they are made on you in the name of the preservation of the species, are unjustified" (54).

What would be important to Steve is the phenomenological presence and motivational efficacy of his affective response (because that his how he consciously accesses emotional information), whether his information is reliable, etc. So, if we want to understand normative moral judgments in terms of affective response, we should not turn to the automatic cognitive appraisals which we can explain in neurological or evolutionary terms, but rather to the phenomenological presence and motivational efficacy that accompany affective responses, as they are the most relevant factors to the individual pondering a decision. As I concluded in the last chapter, these aspects of affective response just are the aspects that are pertinent to an analysis of the normativity of affective responses.[14]

Therefore, if we want to explain the normativity of moral judgments made in terms of affective response, we cannot turn to evolutionary accounts of the reasons why we make the automatic appraisals (which, remember, are also implicated in affective response) that

[14] My original inspiration for these claims comes from an argument made by Christine Korsgaard: "The question how we explain moral judgment is a third-person, theoretical question, a question about why a certain species of intelligent animals behaves in a certain way. The normative question is a first-person question that arises for the moral agent who must do what morality says" (16). While this sounds right, it does seem that third-person accounts are important to normative accounts because they tell us what justifications we should not give for taking something as normative. It excludes us from relying on justifications that are untrue, for example, that the appraisals we make are justified because they are made according to the categorical imperative – assuming that something's falsity is a salient first-person reason not to rely on it as a justification.

we do. Evolutionary accounts of affective response simply do not provide people with good enough reasons to accept their responses as conferring normativity.

Conclusion

The primary function of the first half of this chapter was to work out how it is that people can make moral judgments without feeling, or when they think they have lost any disposition to feeling. I also briefly explained that people who actually have lost all disposition to feel toward something can still make what appear to be moral judgments about that thing, although these judgments would probably be understood by such people more so as conventional rather than moral rules (where this is understood in terms of the conventional/moral psychological distinction).

The second half was mostly dedicated to a close analysis of the relationship between evaluative judgment and affective response through a discussion of the emotional deficit associated with psychopathy. There we found that the emotions cognitivism that Darwall et al. describe in their objection provides a gross oversimplification of the issues, and that we it is probably not appropriate to explain affective responses in terms of evaluative, propositional judgments. In the second half, I discussed the automatic appraisals that cause affective response, and how these appraisals are not like traditional (moral) judgments, and how they provide very little by way of *positive* normative guidance. There we found, as we did in the first chapter, that the phenomenological presence and motivational efficacy associated with affective response are more important to a discussion of whether affective responses can be justified as reasons for normative moral judgment.

Justification, Coherence, and a Modest Theory

In the first chapter, I anticipated that the best method for justifying affective responses as reasons for normative moral judgments would be a suitably precise notion of coherence. This was largely because 'external' methods of justification of affective responses proved to be undesirable, which left us to try an 'internal' method, which is best thought of in terms of coherence. In addition, I argued that the phenomenological presence and motivational efficacy associated with affective responses make them *prima facie* normative, and as such they provide us with good starting points for establishing that certain affective responses can be granted full-blown normativity. The second chapter again established that it is these qualities of affective responses that are relevant to the agent trying to decide which normative moral judgments to make, while also showing that people can profitably think about their affective responses. As will become evident, these conclusions will prove to be crucial to establishing the viability of the following project.

In this chapter, I will attempt to show how the notion of coherence can be used to develop a method for justifying individual affective responses. To do so, I will first closely examine Linda Radzik's paper "A Coherentist Theory of Normative Authority," which will help us better understand how the notion of coherence can perform as a normative function. As will soon become evident, Radzik has done some important groundwork with regard to showing how the notion of coherence can perform a normative role. Once armed with an understanding of how the notion of coherence can be applied to normative question, I will proceed to offer some suggestions on how to justify one’s individual affective responses as reasons for making normative moral judgments.

Normativity and Coherence

Before an attempt to outline how the notion of coherence can be applied to affective responses can be productive, we should first examine Linda Radzik's recent formulation of a coherentist theory of normative authority, with the intent of developing more tools with which to outline a theory for justifying affective responses. Radzik main questions involve normative authority. What, she asks, makes an 'ought' claim authoritative? In other words, what makes an 'ought' claim normatively binding? She believes the answer lies in justification - 'ought' claims have normative authority if they have been justified in a certain way (24). According to Radzik, a justification of this sort must fulfill five conditions, four of which I will discuss.[15] This type of justification must be: (a) first-person accessible, (b) regress-avoiding, (c) reflexive, and (d) comprehensive. I will explain each of these conditions and their importance to any theory of justification in turn. More importantly, once I have gone through Radzik's conditions and have briefly explained her theory, I will also examine how each of these conditions can be applied to our central question about justifying affective responses.[16]

The first condition that Radzik imposes on a normativity conferring justification is that it must be first person accessible. By this she means that the justification of an ought claim had better be convincing from the point of view of the agent who is evaluating this justification. If an agent is supposed do be bound by a normative claim, the reasons for that claim should be both convincing and accessible to that agent. In defense of this

[15] The reason for this is that two of the conditions, first person accessibility and transparency, are so similar that given our purposes we need not distinguish between them. In case the reader is curious about how these conditions interact, Radzik writes this about the transparency condition: "A norm cannot really be justified for an agent if, when she reflects on what it is that makes the norm justified, she can no longer endorse it" (29).

[16] Radzik speaks throughout her paper about coherence between norms. Although her central case is clearly different from ours (i.e., norms vs. affective responses) the theory she develops is amendable to the development of an account involving affective responses.

claim, Radzik writes, "whatever it is that makes morality binding on this agent had better be something she can grasp and appreciate the import of, otherwise her question [e.g., why should I be moral?] has not really been answered". Moreover, she argues that if our theory of justification is to be of any utility, it needs to be first-person accessible: "If thinking about normative authority is supposed to be of any use to me when I am trying to make choices, I had better be able to tell what is authoritative and what is not" (26).

On this matter, I tend to agree with Radzik. Recall that in the last chapter I argued that evolutionary accounts of the implicit appraisals that cause felt affective responses are inadequate as normative justifications precisely because they are not convincing from the first person perspective. In addition, recall that this fact lead me to claim that it is the phenomenological presence and motivational efficacy implicit in affective responses that are relevant to trying to justify them, specifically because they are the elements of affective response that are accessible from the first person perspective. Indeed, it has been an underlying assumption throughout this paper that questions about the justification of affective response must be given an answer that is acceptable from the first person perspective.

My primary reasons for making this assumption stem largely from arguments made by Christine Korsgaard. I agree with Korsgaard that questions about morality are best interpreted as questions asked from the first person perspective, and therefore that they deserve a first-person answer (14-17). She writes, "the question how we explain moral behaviour is a third person, theoretical question, a question about why a certain species of animals behaves in a certain way. The normative question is a first person question that arises for the moral agent who must do what morality says" (16). Unless the reasons why

a person ought to do something are at least somehow accessible to and motivating for that person, at least after they have reflected on and given serious thought to the matter, it is very difficult to see how they could possibly be reasons for that person.[17] There are, of course, several complications involved in holding to this position - for instance, it seems that a person can have a genuine reason for acting of believing without knowing that he has that reason.[18] As can be seen from the discussion in the preceding chapter about the cognitive appraisals that cause felt affective responses, and their lack of normative efficacy from the first person perspective, this seems particularly true when affective responses are under consideration.

This view is motivated by the idea, shared by many ethical theorists, that normativity must somehow be grounded from within the perspectives of individual agents if normativity is going to find a place in the natural order of things (Darwall, 261). This idea often goes by the label 'internalism.' Within the literature, the word 'internalism' has been used to define several positions that ethical philosophers have taken up on issues such as moral judgments, motivation, and reasons.[19] To be clear, the type of internalism that I am talking about in this paper is best understood as being about reasons. It requires that if there really is a reason for an agent to do or believe something, then that agent must be able to appreciate the force of that reason from her own perspective, where it is understood that the agent in question is (a) fully informed of all the relevant facts, and (b) is able to calmly reflect on the matter at hand (e.g., the agent is free of debilitating

[17] Bernard Williams offers strong arguments for thinking this way. See his "Internal and External Reasons." Of course, Williams and Korsgaard disagree quite a bit, but they seem to agree generally on the internalist intuition.

[18] For more on this and related issues, see Peter Railton's "Moral Realism."

[19] For further reading on the differences between various forms of internalism in ethics, see especially Stephen Darwall's "Internalism and Agency" and "Reasons, Motives, and the Demands of Morality."

pathologies or extreme practical irrationality). There is of course much to be said about ethical internalism in general, and the specific form of it that I have just discussed. Nonetheless, I ask the reader to be charitable on this point and to allow me to continue making the assumption that reasons must be in principle accessible to agents from the first person perspective.

The second condition is that our method of justification avoids regress. It is easy to see how regress could infect a justificatory theory, and like Radzik, I believe that avoiding regress is important. For, I could justify my judging that stealing is wrong by saying ‘stealing is wrong because causing other people harm is wrong, and stealing harms other people.’ The next question, though, is whether my judgment that 'causing other people harm is wrong' is itself justified. I could now say that 'harming people causes pain, and causing pain is wrong.' But then I would have to justify this last claim, . . . , and so on. Radzik's suggestion for solving this problem is to apply the notion of coherence to justification. According to her, a norm "will be justified if and only if it coheres well with the norms he accepts" (30). Thinking about justification this way helps avoid regress because "justifying reasons are provided for each norm, though the set of norms is finite and no norm is taken as justified in itself" (30). While I have more to say about avoiding regress, I will reserve those comments for later.

The third condition is comprehensiveness. Radzik notes that there are many different kinds of narrow justification: prudential justification, epistemic justification, moral justification, and so on. She calls these 'interest-driven' types of justification, presumably because each tends to be fairly parochial in its aims, and none of these types of justification seems to capture the notion of normative authority *simpliciter*; indeed,

sometimes one or another of these interest-driven justifications (or, rather, theorists who defend them) might claim normative authority over the others, as morality is often thought to do. To help avoid these problems, she argues, we need a different kind of justification: justification *simpliciter*. She writes:

> If normative authority is really a matter of justification, it must be a very different kind of justification than the interest-driven ones. It must be justification from some more *comprehensive* point of view - a point of view from which we can look over all the interest-defined evaluative schemes and judge which ones should be allowed influence over our choices. Normative authority needs to be characterized in terms of an all-things-considered sort of justification. (27-28).

Clearly, as we are talking about a very narrow form of justification, namely, justifying affective response(s) as reasons for normative moral judgments, we need to see how this narrow from of justification might fare in relation to Radzik's overall scheme. Before we can do that, though, we should first investigate the rest of Radzik's conditions and her coherentist theory of normative authority in more detail.

The fourth condition is reflexivity. If we are going to come up with a theory of justification *simpliciter*, we need to be justified in accepting that that theory of justification itself. Radzik puts it this way: "It must be the case that the alleged standard of justification passes its own test" (31). This condition stems from the comprehensiveness and regress-avoiding conditions. For, if our standard of justification did not pass its own test, it would not be comprehensive, as there would be norms whose justification is left unaccounted for. Also, it would not avoid regress, for we could always

ask if we are justified in accepting certain norms as justified *simpliciter*. Thus, she proposes that a theory of justification *simpliciter* should be self-justifying.

Radzik claims to have developed a theory that meets all these conditions. With these conditions in place, her first step is to define when a norm has been reflectively endorsed: "A norm that is 'reflectively endorsed' is one that is both accepted itself and supported by something else that the agent accepts" (32). One norm supports another when the latter norm satisfies the standard set out by the former norm. For instance, the norm 'do not hit your younger brother' is supported by the norm 'do not harm other people,' because not hitting one's younger brother is a form of not harming other people.

Moving on, she calls her theory Reflective Endorsement Coherentism, and she defines it in the following way:

> **REC**: A norm is decisively justified for an agent if she would endorse it upon reflection and that endorsement would be ultimately undefeated by the rest of her acceptance set. (32)

However, REC is itself in need of justification, so as to satisfy the comprehensiveness condition. Radzik sees RE1 as a kind of restatement of the idea behind REC:

> **RE1**: One ought to accept the norms that one would endorse upon reflection.

Working from this restatement, Radzik claims that an agent could justify RE1 by appeal to RE2, which is quite similar to both REC and RE1:

> **RE2**: I ought to accept the norms of reflection that I do. (34)

So, if I think to myself 'why should I accept the norms that I do?', I can answer my question by reference to RE1: 'One ought to accept the norms that one would endorse

upon reflection.' If I ask myself 'why ought one accept norms that one would reach upon reflection?', I can answer my question by analogy to my own case, which just is a reference to RE2: 'I ought to accept the norms of reflection that I do.' With regard to the reflexivity condition, REC can be restated as RE1, the reason for accepting RE1 is RE2, and the reason for accepting RE2 is RE1 (34). In this way, the theory appears to fulfill the reflexivity condition; along with Radzik, we will refer to it as the 'loop of reflective endorsement.' Each of RE1 and RE2 accounts for the justification of the other. This 'loop of reflective endorsement' also meets the comprehensiveness condition, as it covers all conceivable kinds of interest-driven norms - any norm a person could confidently accept upon reflection passes this test. In addition, it avoids regress - recall that because this theory invokes coherence, each norm is justified by other norms that have been justified upon reflection, the set of norms is finite, and no norm is taken to be justified on its own. Therefore, any query regarding the justification of a norm can be answered by citing other norms that the person in question has accepted, and because that set is finite, eventually the justifications will stop - presumably with the ‘loop of reflective endorsement’ RE1 and RE2.

With regard to the first condition that we discussed, the first-person accessibility requirement, the verdict is mixed. Radzik makes a good point that if an agent does not accept REC, and consequently rejects RE1 and RE2, it seems that she does not trust her own powers of judgment, and therefore ceases to be an agent in some important sense. Indeed, this reasoning does lead one to think that REC must be both be understood and appreciated by any reasonable agent.

However, remember that the purpose of this paper is to inquire whether we can justify

affective responses as reasons for making normative moral judgments. Now, if a person asks the question 'should I accept that affective response provides justifying reasons for holding to the normative moral judgments that I make?', it is hard to see how appealing to REC would be of any help. Indeed, if our worry is narrowly justification of affective response, then simply being told to accept whatever conclusions we reach through reflection does us no good, because doing so begs the question that our reflective methods for evaluating affective responses are indeed reliable and themselves justifiable. Appealing to RE1 or RE2 when asking the question 'does my anger at X provide a justifying reason for judging that X is wrong?' provides little information about whether one is justified in judging that X is wrong. It might help us once we have reached a decision about the matter, but until then, we need first to develop methods for reflecting on affective responses.

Given these considerations, it seems that REC is in some sense lacking by way of comprehensiveness, insofar as it does offer an acceptable account of justification within the narrow, interest-driven types of justification. For our purposes, what this shows is that we need to work out moral justification before we can consider justification *simpliciter*. This next section, then, is dedicated to applying the outline of Radzik's coherenist theory of justification to our more modest goal of showing how affective responses might be given a coherentist form of justification.

Justification, Coherence, and Affective Response

Before developing the outlines of a theory for a coherentist justification of affective responses, let's first take another look at Radzik's conditions, to see which of them should constrain our theorizing. For the sake of exposition, I will begin with those that do not

present major constraints as far as we are concerned, followed by those that should constrain the theory. After the constraining conditions are in place, I will use a few examples in an attempt to show how an affective response could be justified as a reason for a normative moral judgment.

First, with regard to the comprehensiveness condition, because this theory of justification is intended to be a narrowly about affective responses and basing moral judgments on them, it does not seem that it needs to be comprehensive. Subsequently, because the reflexivity condition is largely a result of requiring comprehensiveness, as described above, it will not be a concern either. As I develop the theory, though, I will have something to say about how moral justification relates to other interest-driven kinds of justification, and, if space and time permit, how it relates to justification *simpliciter*.

Second, it is a given that our theory has to be first-person accessible. Indeed, this condition already grounded the reasoning in chapter two that motivated taking the phenomenological presence and motivational efficacy that accompany affective response as the elements of affective response that are relevant to justifying them. Because it is these elements of affective response that are accessible to the first person perspective, it will be these elements that will prove relevant to testing affective responses for coherence.

Third, it does seem that our theory should avoid regress. Given this, however, it also seems that there is a difference between avoiding regress and giving our justifications 'unshakeable foundations,' or even having them terminate in a loop of reflective endorsement, as is the case in Radzik's theory. Now, in the coherentist theory of justification that I will develop, nothing will be beyond question - all conclusions reached

regarding justification will be in principle defeasible. So, while any conclusion reached will always be in danger of being overturned, and thus could be subject to endless demands requiring that we test to see if it still is justified, these are not the types of question that threaten regress. Let me explain.

Recall that I argued in chapter one that affective responses are to a very real extent self-justifying just in virtue of their phenomenological presence and motivational efficacy. This suggests that justifying an affective response is not going to be a matter of saying that it is justified because it has *positive support* by some other affective response that is already been granted 'full' justification. Instead, justifying an affective response will be more a matter of seeing whether any other affective response(s) (each of which, remember, carries its own level of self-justification) conflicts with the phenomenological or, more importantly, motivational tendencies of the original affective response, insofar as they relate to making moral judgments. Justifying an affective response will not require that other, already justified affective responses provide it with positive support. Remember, regress threatens when we justify one thing in terms of another, and then ask what the latter thing is itself justified in terms of, . . . , and so on. Because the theory I will develop below will not justify affective responses 'in terms of' anything else, the regress problem, while certainly a condition we have to fulfill, does not seem to present a pressing concern.

At this point, the best way to proceed is to introduce an example to work through. Recall from chapter two the story about Sarah, who owns a coffee shop. Sarah has recently become conflicted about the positive affective response she feels when she makes a good profit, because she knows that the coffee beans she uses are grown by

Latin American farmers who work as de facto slaves, thereby suffering greatly partly so that she can sell cheap coffee. Knowing that others have suffered partly so that she can sell cheap coffee causes Sarah to have a negative affective response towards the profits she makes - in other words, she feels guilty. On the basis of this feeling, Sarah is tempted to judge that her making the profits that she does is wrong, on account of the fact that it perpetuates the suffering of others. Would she be justified in making this judgment based on her negative affective response to making profits?

In chapter one, I argued that affective responses are largely self-justifying as reasons for making normative moral judgments, simply because of their phenomenological presence and motivational efficacy. However, I also argued that this very quality of affective responses can lead us astray, as when a feeling of disgust causes people to make hasty moral judgments based on unjustified affective responses. As applied to Sarah's case, her just having a negative affective response toward her own profits does provide a *prima facie* justified reason for judging that her making profits is wrong. However, she should not make a hasty judgment. The first thing she should do before beginning to evaluate whether her affective response is justified is calm herself. She first needs to get to the point where she is not overrun by emotion by making a conscious effort to regulate her response. Therefore, it seems that the first thing that a person needs to do when attempting to justify an affective response, then, is to regulate her emotional state. In Pierre Phillipot's terms (see chapter two, section two), in doing this one would be regulating one's 'schema' activation so that one only gets what emotional information one thinks one needs from it.

Once Sarah is calm, she needs to see if there are any other affective responses that she

might have that 'conflict' with her original affective response. Needless to say, before we can give Sarah any advice, we need an explanation of how two affective responses could 'conflict' with each other, or, in other words, how they could cohere or incohere with each other. At first glance, affective responses taken alone do not seem to cohere or incohere with each other. If I am happy at getting accepted to the graduate school of my choice, and subsequently angry at my best friend, those responses do not seem to cohere or incohere. Feeling happy at one point and angry at another does not seem to entail that there is any coherence-type relationship between the two emotions. It seems that affective responses need a common object to cohere or incohere, i.e., if I felt both happy and sad toward the same object. The way to get around this is to say that if an agent has two or more discrete affective responses caused by the same stimulus *or* sufficiently similar stimuli, that person has potentially conflicting affective responses toward that stimulus(i).

With this in place, I propose that for the moment we understand both coherence and incoherence relationships between the elements defined above through phenomenological feel and motivational efficacy. Taking these aspects of affective response to be the ones that are relevant to justification is in line with our first-person accessibility requirement. With regard to the different phenomenological 'feelings' associated with different affective responses, and how they can cohere or not, I will talk about affective responses being phenomenologically 'positive' or 'negative' in relation to their object(s). For example, feeling happy at some object would be positive, and feeling disgusted or angry at the same object would be negative. Let's presume that if two affective responses are both either phenomenologically 'positive' or 'negative' they *prima facie* cohere, and if one is 'positive' and the other 'negative' they *prima facie* incohere.

With regard to motivational efficacy, I mean narrowly the motivational push that an affective response might have that would terminate in a person to rendering a certain moral judgment. Once again, we will talk about affective responses being 'positive' or 'negative.' For instance, when a person feels guilty because he has stolen from his friend, his guilt would motivate him to judge that his actions were morally wrong. An affective response like this would be considered 'negative.' On the other hand, if a person feels happy that another person is helping alleviate the suffering of others, that person would be motivated to judge that that person's actions are morally right. An affective response like this would be considered 'positive.'

Returning to Sarah's case, it seems that she does have conflicting responses. For, although she feels guilty about making a profit, she also feels happy at making a profit. Given that the phenomenological feels associated with these two emotions conflict it would seems that these two responses are incoherent, from which it should follow that Sarah's guilt is not justified as a reason for making a moral judgment. However, do they conflict motivationally, in the sense defined above? In other words, should Sarah's happiness recommend bear any relevance to making a moral judgment?

Developing a principled method for distinguishing between morally relevant and irrelevant emotions in all cases would not be an easy task, and it is not one that I will endeavor to accomplish in full. Nonetheless, I think that we can distinguish between morally relevant and irrelevant emotions at least one important sense, which will at least allow us to deal with Sarah's case.

In chapter one, I discussed the psychological distinction between moral and conventional norms while describing Shaun Nichols' argument against moral objectivity.

While in that discussion I emphasized the difference between this distinction in terms of authority contingency because of it's obvious bearing on moral objectivity, here I will emphasize another difference: the presence or absence of victims. Moral transgressions, as defined by the moral/conventional psychological distinction, always involve harming a person in some way, whereas conventional transgressions involve violating the social order by departing from the standard behavioural patters that structure social interactions (Blair, "Neuro-Cognitive Systems" 14). Again, I do not want to make the claim that, all psychological distinctions aside, all norms that a person could properly call *moral* norms necessarily involve prohibiting harming other people in some way (i.e., there might be moral norms about harming animals, or about promoting the welfare of others). However, I think it is at least fair to say that a very important subset does in fact prohibit harming others, and for the sake of simplicity and getting to the point, I will work only from within this subset.

With this in mind, I propose that for the remainder of this paper we understand an emotion to be relevant to rendering moral judgments if it directly involves harming others or norms that prohibiting harming others, where 'harming others' indicates causing someone either physical or substantial economic harm that results in greatly reduced well-being of the victim (there are, of course, other kinds of harm, e.g., psychological harm). Thus, guilt is a relevant emotion if a person feels that way for harming another person; happiness is a relevant emotion if a person is happy because a stranger stopped a thug from stealing her purse, and so on.

Let's return once more to Sarah's case. She has two discrete feelings toward her making profits: she feels guilt on the one hand, and happiness on the other. Her guilt is

clearly a morally relevant emotion, as she feels guilty specifically because Latin American farmers suffer greatly partly so that she can buy cheap coffee. Does her happiness pass the same test? Well, let's say that she feels happy because making money allows her to live well and save for early retirement.[20] In this case, her emotion is not morally relevant, as it does not relate to harming others, nor does it relate to norms against harming others. Indeed, she seems to be having a positive emotional response to making money specifically because it is in her self-interest.

So, it would seem that Sarah should come to the conclusion that her guilt is justified as a reason for judging that her making profits is wrong, because her happiness is not motivationally relevant to making moral judgments. Now that Sarah has reached a conclusion about whether she is justified or not, perhaps we can formalize the method whereby she reached that conclusion.

To do so, I will discuss and reformulate two of Radzik's more important statements regarding her theory of Reflective Endorsement Coherentism. She makes two main claims: one that tells us when a norm is reflectively endorsed, and another that tells us when it is justified. First, "a norm that is 'reflectively endorsed' is one that is both accepted itself and supported by something else that they agent accepts" (32). Second, "a norm is decisively justified for an agent if she would endorse it upon reflection and that endorsement would be ultimately undefeated by the rest of her acceptance set" (32). Let's apply these two claims to our case.

With regard to 'reflective endorsement,' there are two pertinent aspects of Radzik's statement. First, there is the issue of an affective response being supported by another as

[20] In addition, let's assume that she has other employment options, does not support a family, and would not suffer an inordinate amount by changing her business practices, etc.

a condition for receiving endorsement. In chapters one and two, I argued that our affective responses are *prima facie* self-justifying just in virtue of their phenomenological presence and motivational efficacy. In addition, in my discussion of regress, I argued that this leads to the conclusion that we do not *need* to justify our affective responses in terms of *anything else*, just because affective responses just are *prima facie* normatively compelling. Therefore, we do not need our affective responses to be supported positively by any other already justified responses; rather, what we are looking for in testing for coherence in this context is testing for whether there are any reasons *not* to make a judgment based on a certain affective response. This means that the type of coherence we are talking about is not the broad type that is associated with, for example, the Rawlsian idea of 'wide reflective equilibrium,' but rather is a very narrow and localized conception of coherence. Second, there is the issue of something's being 'accepted itself.' Again, relying on the *prima facie* self-justifying and normative nature of affective responses, we can say that affective responses come with a certain level of acceptance.[21]

With regard to Radzik's formal statement of REC, we can now reformulate it for our purposes as the Formula for the Justification of Affective Responses (FJAR):

> An affective response is decisively justified as a reason for making a normative moral judgment for an agent if there are no other motivationally relevant affective responses that conflict (incohere) with it, where 'motivationally relevant' means that the response in question motivates the agent to make a conflicting moral judgment.

This principle is fairly simple: if you initially have an affective response that suggests

[21] To make an analogy, this level of acceptance on its own is something like the data priority that coherentist theories of empirical justification theories often invoke; see for example Thagard's "Coherence: the Price is Right."

making a moral judgment, unless you have another affective response that recommends making a conflicting moral judgment, then the original response is justified as a normative basis for making the moral judgment that it recommends.

An obvious objection looms at this point: FJAR cannot resolve motivational conflicts that do arise between affective responses. For instance, if a person does have motivationally conflicting affective responses, FJAR only tells her that she should not render a normative moral judgment; it gives no advice on how to resolve this conflict.

In lieu of resolving this problem, I will only tease out one possible response. We could reply by developing a hierarchical system that would rank the types of motivationally relevant affective responses as more or less motivationally relevant, which could help to resolve many conflicts. In this system, for instance, affective responses that are caused by the distress of another person could 'trump' affective responses that would motivate one to judge that it is permissible to harm someone in order to promote the welfare of another person. Thus, for example, if one was conflicted about whether to steal a substantial amount from another person in order to send one's child to university, one would come to the conclusion that doing so would be wrong. However, a methodology like this would not be able to resolve conflicts between affective responses of the same type, e.g. between two affective responses that are caused by the distress of others. In this case, the only advice I can currently conjure up would be to 'trust your gut,' or simply to withhold judgment. Developing this hierarchical system could help with some cases, but again, in many difficult circumstances, the advice it would give would not be very compelling.

Despite the current shortcomings of FJAR, it does provide justification for many of

the things that all people judge to be wrong: we are all justified in judging that murder, exploitation, assault, theft, etc., are wrong, extenuating circumstances notwithstanding. In addition, with regard to more difficult cases, it is certainly plausible that FJAR could be expanded to offer more specific advice on how to resolve conflicts, and the reply I outline above provides something of a start.

Looking Forward: FJAR, REC, and Moral Reasoning

In the following, I will briefly discuss how FJAR relates to Radzik's REC and how it could function in relation to moral reasoning.

Recall that one condition that Radzik places on REC is that it be comprehensive. A comprehensive theory is supposed to give us "a point of view from which we can look over all the interest-defined evaluative schemes and judge which ones should be allowed influence over our choices" (32). As Radzik's REC stands, it does not give us much advice about to what degree the norms associated with normative moral judgments should override other norms, such as prudential norms. As REC and FJAR are both subjectivist theories, a definite answer that applies to each and every individual is likely not forthcoming. Instead, we could develop REC so that it offers individuals advice about when they should take moral norms to be overriding and when they should not. For instance, because people often engage in various forms of self-deception to justify their acting in ways that, if they were not motivated otherwise, they would judge to be immoral, we could recommend methods and patters of thinking that help people avoid this variety of self-deception.[22] The idea behind this would not be to develop a definitive formula that an individual could use to determine whether demands of morality are

[22] For more on this variety of self deception, see Lorraine Besser-Jones' "Social Psychology, Moral Character, and Moral Fallibility."

overriding in any given situation. Rather, we would be suggesting ways whereby a person could himself decide which norms he *really* thinks are overriding, or which norms he would decide really were overriding in the absence of distorting factors, such as self-deception. In fact, I have already given some advice on this matter by distinguishing between morally relevant and irrelevant affective responses. Clearly, though, more work on this issue needs to be done.

Moving on, when people engage in moral reasoning, they tend to do so by working from general principles to particular judgments, through analogies, by deliberation, etc. For example, it seems perfectly reasonable for a person to work from the general principle that causing other people pain is wrong to the particular judgment that beating up one's younger brother is wrong (deduction), or from the judgment that picking on people at school is wrong, so picking on people at cub scouts is wrong. In each of these cases, the concluding judgment seems to be valid; the moral judgment that we reach in either case seems justified, even though FJAR has not been applied.

To account for this, I suggest that for an instance of moral reasoning to be valid, it needs to *preserve* justification, or simply produce a justifiable conclusion, where justification is understood through FJAR. This means that any concluding judgment reached through moral reasoning has to pass the same test that affective responses have to pass. In other words, the concluding judgment cannot be in conflict with any other motivationally relevant affective responses. In the above analogical case, justification is preserved because the situations involved are sufficiently similar - in moving from the one scenario to the other, no new relevant affective responses would be triggered, and thus the concluding judgment is justified. In the deductive case, the situation is a little

different. For, it involves the *general* principle that causing other people pain is wrong. Nonetheless, a general principle like this only confers justification, and is itself justified, one would think, if specific instances of causing people pain prove to be wrong. So, general principles only confer justification insofar as the specific situations that they describe tend to prove to be either wrong or right, whichever the case may require. In the case of the principle that causing people pain is wrong, it will probably turn out that in the majority of instances where one person causes another pain, we are indeed justified in judging that action to be wrong. Therefore, the principle 'causing other people pain is wrong' is likely a fairly reliable justification conferring principle.[23]

To make clear what I mean here, I will consider an instance where the conclusions that we reach through moral reasoning do not turn out to have full-blown justification on an application of FJAR. First, consider the principle 'stealing is wrong.' Now, consider the specific judgment that we can draw from this principle: 'stealing from the grocery store in order to feed one's family is wrong.' Here the concluding judgment is at the very least questionable. For, although people certainly tend to be motivated to judge that stealing is wrong because they have a negative affective response to stealing (e.g., anger), simply working from the general principle does not establish itself that this specific implementation of it is valid. For, although people tend to be motivated to judge that stealing is wrong, people also tend to be motivated to judge that feeding one's family is right, and many people presumably are motivated to judge that stealing to feed one's family is right. To be clear, I am not saying that the judgment 'stealing to feed one's family is right' is necessarily justified, but only to show that just working from a general

[23] In fact, there might be general principles that are always justification conferring: for instance, the more specific judgments/conclusions that people reach from the principle that 'wanton cruelty to other people is wrong' are unlikely to be overturned by an application of FJAR.

moral principle does not ensure that the conclusion is justified.

With regard to both REC and moral reasoning, questions obviously remain, and they certainly are worthy of investigation. However, the goal of this paper was only to show, in outline, how a coherentist theory of justification could be applied to affective responses and, subsequently, the normative moral judgments we base on them. Given this modest goal, I think I have shown that a more developed project of this nature could prove profitable.

Conclusion

Throughout this paper, my explicit goal was to show how a person could go about justifying his affective responses as reasons for making normative moral judgments. In reaching this goal, the most important conclusions that I argued for were that normativity is *prima facie* inherent in affective response, that affective response does indeed play an important explanatory and causal role with regard to an emotion-heavy sentimentalist theory of moral judgment, and that it is at least possible for a person to go about justifying his moral judgments by considering his affective responses.

The underlying motivation for this paper, though, is more general. I began with a quote from Peter Railton, and it is worth reading again:

> Ethical Naturalism has yet to find a plausible synthesis of the empirical and the normative: the more it has given itself over to descriptive accounts of the origin of norms, the less has it retained recognizably normative force; the more it has undertaken to provide a recognizable basis for moral criticism of reconstruction, the less has it retained a firm connection with descriptive social or psychological theory.

Each of the more important conclusions that I argued for in this paper were intended to help resolve this conflict. However, one conclusion in particular could go far in bringing together descriptive and normative moral discourse.

With regard to the *prima facie* normativity inherent of affective response, in a naturalist theory, as Stephen Darwall notes, the only thing that normative force can be is motivational force (Internalism 168). Because affective response is so closely tied to motivation, we can plausibly claim, within a naturalistic scheme, that affective response

is intimately connected with normativity. This seems to be an especially potent combination for sentimentalists with a naturalistic bent, as it provides them with a potentially very profitable starting point from which to engage in normative moralizing. Indeed, taking this tack helps resolve some of the more salient problems associated with developing a normative theory, namely, first-person accessibility and regress.

Nonetheless, this approach to normativity does give up on many of the things that several theorists believe are necessary for a normative theory to be valid, namely, moral objectivity ('external' justification of affective responses), and an unshakeable foundation from which we can claim that the norms of morality decisively override other norms, such as norms of prudence. However, if sentimentalists are prepared to give up these aspects of traditional normative theory, they have at their disposal a very naturalistic basis from which to develop a normative moral theory.

Bibliography

Baron-Cohen, Simon, S. Wheelwright, J. Hill, Y. Raste, and I. Plumb. "The 'Reading the Mind in the Eyes' Test Revised Version." Journal of Child Psychology and Psychiatry. 42 (2001): 241-251

Bechara, Antoine. "A Neural View of the Regulation of Complex Cognitive Functions by Emotion." The Regulation of Emotion. Ed. Pierre Philippot and Robert S. Feldman. New Jersey: Lawrence Erlbaum, 2004

Besser-Jones, Lorraine. "The Empirical Foundations of Moral Behaviour." under review.

Blair, Irene V., Charles M. Judd, and Jennifer L. Fallman. "The Automaticity of Race and Afrocentric Facial Features in Social Judgments." Journal of Personality and Social Psychology 87 (2004): 763-778

Blair, James. "A Cognitive Developmental Approach to Morality: Investigating the Psychopath." Cognition. 57 (1995): 1-29

Blair, James, L. Jones, F. Clark, and M. Smith. "The Psychopathic Individual: a Lack of Reponsiveness to Distress Cues?" Psychophysiology. 34 (1997): 192-198

Blair, James, E. Colledge, L. Murray, and D. Mitchell. "A Selective Impairment in the Processing of Sad and Fearful Expressions in Children with Psychopathic Tendencies." Journal of Abnormal Child Psychology. 29 (2001): 491-498

Blair, James, D. Mitchell, K. Peschardt, E. Colledge, R. Leonard, J. Shine, L. Murray, and D. Perrett. "Reduced Sensitivity to Others' Fearful Expressions in Psychopathic Individuals." Personality and Individual Differences. 37 (2004): 1111-1122

Blair, James, Derek Mitchell, and Karina Blair. The Psychopath: Emotion and the Brain. Malden: Blackwell, 2005

Blair, James, A.A. Marsh, E. Finger, K.S. Blair and J. Luo. "Neuro-Cognitive Systems Involved in Morality." Philosophical Explorations. 9 (2006): 29-43

Blackburn, Simon. "How to be an Ethical Antirealist." Moral Discourse and Practice. Ed. Stephen Darwall, Alan Gibbard, and Peter Railton. New York: Oxford UP, 1997

D'Arms, Justin, and Daniel Jacobson. "Sentiment and Value." Ethics. 110 (2000): 722-748

D'Arms, Justin, and Daniel Jacobson. "The Moralistic Fallacy." Philosophy and Phenomenological Research. 61 (2000): 65-90

Damasio, Antonio R. Descartes Error. New York: Putnam, 1994

Darwall, Stephen, Alan Gibbard, and Peter Railton. "Toward *Fin de Siecle* Ethics: Some Trends." Moral Discourse and Practice. Ed. Stephen Darwall, Alan Gibbard, and Peter Railton. New York: Oxford UP, 1997

Darwall, Stephen. "Reasons, Motives, and the Demands of Morality: An Introduction." Moral Discourse and Practice. Ed. Stephen Darwall, Alan Gibbard, and Peter Railton. New York: Oxford UP, 1997

---. "Autonomist Internalism and the Justification of Morals." Nous 24 (1990): 257-267

---. "Internalism and Agency." Philosophical Perspectives 6 (1992): 155-167

Flanagan, Owen. Varieties of Moral Personality. Cambridge: Harvard UP, 1991

Greenspan, Patricia. “Emotion and Rationality.” The Oxford Handbook of Rationality. New York: Oxford UP, 2004

Haidt, Jonathan. "The Emotional Dog and its Rational Tail: a Social Intuitionist Approach to Moral Judgment." Psychological Review 108 (2000): 814-834

Jones, Karen. "Metaethics and Emotions Research: a Response to Prinz." Philosophical Explorations 9 (2006): 45-53

Korsgaard, Christine M. The Sources of Normativity. New York: U of Cambridge P, 1996

Kosson, David S., Yana Suchy, Andrew R. Mayer, and John Libby. "Facial Affect Recognition in Criminal Psychopaths." Emotion. 2 (2002): 398-411

Lutz, Antoine, Laurence L. Greischar, Nancy B. Rawlings, Matthieu Ricard, and Richard J. Davidson. "Long-term Meditators Self-induce high-amplitude Gamma Synchrony During Mental Practice." pnas.com. 2006. Proceedings of the National Academy of the Sciences of the United States of America. June 15th 2006. <http://www.pnas.org/cgi/content/abstract/101/46/16369>

Marino, Patricia. “Expressivism, Deflationism, and Correspondence.” Journal of Moral Philosophy. 2 (2005): 171-191

Nichols, Shaun. Sentimental Rules. New York: Oxford UP, 2004

Philippot, Pierre, Celine Baeyens, Celine Douilliez, and Benjamin Francart. "Cognitive Regulation of Emotion: Application to Clinical Disorders." The Regulation of Emotion. Ed. Pierre Philippot and Robert S. Feldman. New Jersey: Lawrence Erlbaum, 2004

Prinz, Jesse. "The Emotional Basis of Moral Judgments." Philosophical Explorations 9 (2006): 29-43

Radzik, Linda. "A Coherentist Theory of Normative Authority." The Journal of Ethics 6 (2002): 21-42

Railton, Peter. "Moral Realism." Moral Discourse and Practice. Ed. Stephen Darwall, Alan Gibbard, and Peter Railton. New York: Oxford UP, 1997

Richell, R.A., D. Mitchell, C. Newman, A. Leonard, S. Baron-Cohen, and J. Blair. "Theory of Mind and Psychopathy: Can Psychopathic Individuals Read the 'Language of the Eyes?" Neuropsychologia. 41 (2003): 523-526

Sabini, John, and Maury Silver. Moralities of Everyday Life. New York: Oxford UP, 1982

Schroeder, Tim. Three Faces of Desire. New York: Oxford UP, 2004

Stein, Richard, and Carol Nemeroff. "Moral Overtones of Food: Moral Judgments of Others Based on What They Eat." Personality and Social Psychology Bulletin 21 (1995): 480-490

Thagard, Paul. "Desires are not Propositional Attitudes." Dialogue: Canadian Philosophical Review.

Thagard, Paul. "The Emotional Coherence of Religion." Journal of Cognition and Culture. 5 (2005): 58-74

Thagard, Paul. "Ethical Coherence." Philosophical Psychology. 11 (1998): 405-422

Wheatley, T., and Jonathan Haidt. "Hypnotically Induced Disgust Makes Moral Judgments More Severe." Psychological Science 16: 780-784

Williams, Bernard. "Internal and External Reasons." Moral Luck. New York: Cambridge UP, 1981

Williams, Bernard. "Persons, Character, and Morality." Moral Luck. New York: Cambridge UP, 1981

Zhu, Jing, and Paul Thagard. "Emotion and Action." Philosophical Psychology: 15 (2002): 19-36