Philosophical Issues, 15, Normativity, 2005

TWO ARGUMENTS FOR SENTIMENTALISM

Justin D'Arms
Ohio State University

'Sentimentalism' is an old-fashioned name for the philosophical suggestion that moral or evaluative concepts or properties depend somehow upon human sentiments. This general idea has proven attractive to a number of contemporary philosophers with little else in common. Yet most sentimentalists say very little about the nature of the sentiments to which they appeal, and many seem prepared to enlist almost any object-directed pleasant or unpleasant state of mind as a sentiment. Furthermore, because battles between sentimentalism and its rivals have tended to be joined over large issues about realism and antirealism, or cognitivism and noncognitivism, some attractive reasons for adopting sentimentalism which are to some extent independent of these issues have been largely ignored in metaethical discussion.

This paper aims to motivate sentimentalism, but also to circumscribe its ambitions by rendering explicit some tacit assumptions in moral psychology on which I think the most promising sentimentalism depends. I begin (in section one) by sketching the kind of sentimentalism that I want to defend. Then, in sections two and three, I articulate two positive arguments for a sentimentalist understanding of certain evaluative concepts. The arguments I consider have their origins in the writings of various other authors, I think, but neither they nor their consequences have been clearly articulated before. In section four, I explore just what the sentiments would have to be like in order to play the role required of them in the arguments I develop. I will suggest that these arguments supply a highly specific 'job description' for the states to which sentimentalism appeals. Hence, sentimentalists who want to use these arguments, or ones like them, cannot be as casual about what they mean by 'sentiments' as many have tended to be. I then investigate a category of 'natural emotions' that meets that job description rather nicely, and offer some reasons for doubting that more inclusive categories of

sentiments fare as well. It is a contested empirical question whether there are any natural emotions in my sense, and, if so, which ones they are. So one interesting consequence of this discussion is that it connects the prospects for sentimentalism with some empirical questions in moral psychology.

I. Rational Sentimentalism and Regulative Concepts

Sentimentalists hold, of some range of evaluative concepts, that any adequate account of their content must make some appeal to sentiments or emotions. I call the version of the theory I favor *rational sentimentalism*, and I have been working toward a systematic statement and defense of it in collaboration with Daniel Jacobson. Rational sentimentalism holds that certain evaluative concepts are *regulative concepts* for paired emotion types: concepts whose primary function is to guide or regulate specific kinds of emotional response by appeal to reasons of a particular sort. I'll leave open for now how many evaluative concepts the doctrine is applied to, and which ones.

Let us start with a family of concepts that seems to carry an especially intimate tie to human emotional responses. Some examples are shameful, fearsome, enviable, disgusting, funny, and pitiful. These concepts are evaluative: to apply one to something is to think it good or bad in some way. Each of them is affiliated with a specific and familiar emotional response (in many cases the term expressing the concept is a cognate of the name for such a response). Moreover, judgments applying them invoke reasons: reasons to feel, at least, and perhaps to act as well. Thus, if one thinks a trait shameful, for instance, one thinks its bearer has a reason to be ashamed of it, and one may well think he has a reason to eliminate or conceal it too.

These concepts are promising candidates for some kind of sentimentalist treatment because it is intuitively plausible to suppose that judgments concerning what is funny, shameful, etc., depend upon and engage with our reactions of amusement and shame.[1] But what is the nature of this dependence or engagement?[2] The simplest answer would be that to judge that one of these concepts applies just is to feel the associated emotion, but that is clearly wrong. I can think something funny, disgusting, or shameful without actually having the relevant response. And I can find myself amused by something I judge to be juvenile, and not funny, or ashamed of something I insist is no true blemish. So actually feeling an emotion on an occasion is neither necessary nor sufficient for making the associated judgment on that occasion.

A familiar strategy for circumventing such problems in other cases is dispositionalism. Perhaps, then, these evaluative judgments are about the dispositions of objects to elicit particular sentiments. It is plausible to treat redness as a dispositional concept, imputing a propensity to produce certain

sensations under certain conditions. One might hope for an analogous treatment of concepts such as 'shameful' and 'funny,' on which to think something funny would be to think that normal persons under normal circumstances would be amused by it. However, the dominant contemporary sentimentalist approach to the relationship between sentiments and evaluative concepts has not been dispositionalism, but a second-order sentimentalism, holding that to apply a response-dependent concept Φ to an object X (i.e. to think that X is Φ) is to think it *appropriate* (merited, rational, justified, warranted) to feel an associated sentiment F toward X. In recent years, Elizabeth Anderson, Simon Blackburn, Allan Gibbard, Bennett Helm, John McDowell, Kevin Mulligan and David Wiggins (as well as Jacobson and I) have each attempted to explicate at least some evaluative concepts along these lines. Helm (2001), McDowell (1997a, 1997b), Mulligan (1998) and Wiggins (1987) apparently regard the second-order sentimentalist schema as applying to evaluative concepts quite generally. Blackburn (1993, 1998) and Anderson (1993) appeal to the more inclusive category of 'attitudes' in most contexts, but each treats sentiments as at least one central kind of attitude that can be invoked to explain evaluative judgment. Gibbard (1990) deploys the schema explicitly only with respect to certain examples: shameful, dangerous, and wrong, the last of which receives most of his attention. The schema is second-order because it understands the direct evaluative judgment about X in terms of a normative judgment of 'appropriateness' about our own emotional states, namely, about the having of a sentiment F toward X.

These philosophers don't explicitly characterize their views as 'sentimentalist,' and their theories are quite different in other metaethical respects. One important difference is in how they understand the judgment that a sentiment is appropriate. Blackburn and Gibbard hold, and the others will deny, that this second-order judgment is itself an expression or projection of a further noncognitive state of mind. But all parties here share a commitment to the response-dependency of these concepts, and to a second-order explication of this dependence. The central idea of contemporary sentimentalism is that to judge that one of these evaluative response-dependent concepts applies is not to feel F, nor be disposed to feel it, but to favor feeling F, or to think there is reason to feel it, in response to X.

Jacobson and I have argued that this articulation of the relation between sentiment and value is still unsatisfactory as it stands, because there are too many good reasons for (not) feeling an emotion—too many senses in which it might be (in)appropriate—and many of them have nothing to do with thinking its object is Φ.[3] For instance, it might be inappropriate, because mean-spirited, to envy your friend's well-deserved success. But to think so is surely not yet to deny that the success is enviable. Similarly, one might favor never feeling shame on the grounds that shame makes it harder, not easier, to mend the inadequacies of which one is ashamed. But

this is not to deny that the inadequacies are shameful. These distinctions seem obvious, once made, but their importance for sentimentalism has not been widely enough appreciated.

What is needed here is the notion of a kind of appropriateness that restricts the range of considerations about whether to feel F to just those that speak to whether or not the circumstance is Φ. Call the question of whether F is appropriate in this sense the question of whether it 'fits' the circumstances. Sentimentalists owe some kind of account of how to distinguish considerations of fittingness from other reasons to feel. The most promising place to look, if Jacobson and I are right, is at the emotions themselves. Thoughtful examination of the nature of particular emotions allows one to articulate, in a rough and ready way, what the characteristic concern of a given emotion type is. In effect, one can treat emotions as species of evaluative experience, in the course of which things are presented as mattering in certain specific ways.

Reasons to think that the circumstances are indeed the way a particular emotion 'presents' them as being will be reasons that bear on whether that emotion fits. The constraints on what reasons are relevant to whether something is shameful, fearsome, enviable, etc., are therefore partly determined by features of these emotions themselves. (It is *because* it is not part of the nature of envy to present its object as undeserved that the fact that your friend deserves his success is irrelevant to whether it is enviable.)

While assessments of fittingness are dependent in certain ways on interpreting the actual nature of our involuntary emotional responses (what it is that they are concerned with, and in what way), it is always possible to criticize any particular response as unfitting. Rational sentimentalism's central commitment is to the role of the judgment that an emotion fits in regulating our tendencies to feel and act. That is, thinking an emotion fits is a kind of *rational endorsement* of having the emotion—it must be seen as counting in favor of having the emotion in just the way that the thought that the circumstance is shameful counts in favor of being ashamed of it. Fully vindicating this normative aspect of judgments of fittingness in the face of the constraints above is a daunting philosophical task, however. So while the sentimentalist approach seems a promising one for explicating at least the concepts we've been discussing so far, there is much work still to be done.

In view of these complications, it's worth asking why sentimentalism is led in this direction. What is the difference between these regulative concepts and color concepts, say, which makes the more familiar and simpler dispositionalist conception of response-dependency available for colors but not for values? There are many important differences, I believe, but I will focus on a few that have not been widely noted already. First, color concepts are not properly used to criticize conventional patterns of reaction. We find it useful to categorize things as 'red' or not because there is sufficient uniformity in our reactions that we can expect the great majority of people

to experience the things we call red in the same way—or at least to behave as though they do. We exploit this uniformity in order to pick out which chair, or car, we are speaking of—that is, in order to sort the objects we see in mutually convenient ways. If color responses were substantially less uniform, there would be no obvious point in treating color predicates as referring to context-invariant features of objects, in the way that our actual property-ascribing discourse seems to do. Our color concepts owe their existence to the contingent fact that most people tend to respond similarly to the spectral reflectance characteristics of various surfaces under various circumstances—and the content of these concepts, I claim, reflects this fact. Secondly, we do not deliberate (much) over whether to see things as red, and reflection on failures of our reactions to track an object's true color does not tend, even in the long run, to alter the way it looks to us. Finally, color experience does not have any intrinsic connection to motivation, and color concepts do not have any intrinsic tie to reasons for action.

Contrast these points with an evaluative case. Convince me that most gay teenagers are ashamed of their sexuality and you won't have convinced me that it's shameful.[4] In rejecting common opinion I might be thought by some to betray a defective sensibility, but no one will think me incompetent with the concept 'shameful.' Perhaps because these responses and the concepts to which they give rise matter to us, there is a point in defending the appropriateness of one's own responses against majority reaction—and we recognize such defenses as coherent (though, perhaps, incorrect) even when we side with the majority. Furthermore, although our emotional reactions sometimes persist in spite of the thought that they are inappropriate, such thoughts often alter the way we feel. Sometimes this happens immediately (as when I find out the lab gave me the wrong test results, and I'm actually fine) and sometimes only in the long run (as when you gradually stop feeling threatened by a lover's old friends). Similarly, the conviction that a given circumstance merits a given emotional response that we do not tend to give it has some tendency, over time, to dispose us to feel the way we think appropriate. And, of course, emotions involve motivations in a way that colors do not. So deliberation over how to respond has a special kind of practical purpose in the emotional case that it doesn't in the color case. While these brief remarks cannot establish that the differences between values and colors are great enough to scuttle all dispositionalist accounts of concepts like 'shameful', perhaps they suffice to motivate the consideration of an alternative approach.

II. The Regulative Role Argument

It must be granted that the sentimentalist proposal cannot claim to capture every aspect of ordinary practice—even with respect to those concepts to which it is well suited. But neither can any rival account. The evaluative

concepts we acquire as English speakers likely embody some commitments or assumptions which will ultimately prove unsustainable or mutually incompatible. Recognition of this truism, together with more general doubts about the possibility of identifying analytic truths in any realm of discourse, have changed the aspirations of metaethics over the last thirty years. Contemporary philosophers' accounts of the meaning of ethical terms or the nature of ethical concepts do not typically pretend to accommodate every feature of such concepts.[5] Instead, we aspire to elucidate them in ways that make sense of as much as possible of their use while fitting as well as possible with the rest of what we know about ourselves and our world. In assessing an explication of a certain evaluative concept we should ask to what extent, if we accepted it, we could carry on using the concept in the same inferential, judgmental and practical roles it currently plays for us. Insofar as the account would have us revise these roles, are these revisions that we can embrace and see as motivated by something we took to be important to the practices already? (If not, this proposal may be better placed as an alternative to our present concept than as an explication of it.)[6]

These suggestions are commonplaces of contemporary moral philosophy. It is worth adding another set of considerations, which are not so commonly considered. One can also ask whether a given account of some evaluative concept helps us to see why we have such a concept in the first place, and why the concept has had such longevity as it has. What are the enduring human needs and interests to which it answers? These sorts of questions may lead us to try to articulate a clear *function* for the concept, and show that this is a function that we need some concept or other to serve. If we can show this, I believe that helps to motivate our interpretation of existing terms as expressions of the posited concepts. Of course, the terms must be sufficiently proximate to render the interpretation plausible on more familiar interpretive grounds. But if they are, then such arguments from functional role can be adduced to select between rival interpretations.

It is in the spirit of thoughts like these that I offer what I'll call the *regulative role argument* in support of a sentimentalist account of the concepts under discussion.

1. Human beings are prone to experience emotional responses that are instances of various emotion types.
2. These responses have various features that both incline us and give us reason to attempt to regulate them: that is, to reflect upon, confer about, and develop (more or less articulate) standards concerning when to have them, and to take such steps as we can to feel emotions in accordance with our conclusions.
3. Such reflection and discussion would be well served by a vocabulary of terms that characterized circumstances in terms of their fit with precisely these types of emotional responses.

4. 1–3 supply a role for a vocabulary of terms that stand for concepts that are assessments of the fittingness of our emotional reactions.
5. On the sentimentalist interpretation, various evaluative terms such as shameful, pitiful, and enviable would be just such terms.
6. Therefore, let us take those terms to express regulative evaluative concepts, construed along the lines of the sentimentalist interpretation.

The rest of this section attempts to flesh out this highly schematic argument. First I show how such indirect considerations can count in favor of a particular interpretation of some term. Then, I explain the features of emotional experience adverted to in premise 2 and show how they supply a need to regulate emotional responses. Finally, I say why a vocabulary of terms that served to express fittingness judgments would be especially useful for the reflective and discursive purposes I have in mind.

The argument is indirect, and it does not aspire to be a deductive argument for the correctness of the sentimentalist interpretation of any concept whatever. I tried to signal as much by putting its conclusion in the imperative mood. What I hope the argument can accomplish is to supply reasons for favoring the sentimentalist interpretation of a given range of terms over viable competitors, in a context where the sentimentalist interpretation has some prima facie plausibility in any case, and where no interpretation can claim to capture every feature of the discourse.

Consider, for instance, how one might adjudicate between rational sentimentalism and a form of dispositionalism not addressed above. A speaker-relative dispositionalist might hold that to judge something shameful, funny, etc., is to judge that one is disposed to feel the relevant sentiment, under some relevant set of circumstances. On that proposal, to call something shameful is to say that it has a disposition to make one ashamed of it in oneself, or contemptuous of it in others, under some canonical conditions. Let us grant that this proposal may capture the speaker's intended meaning for some uses of the term 'shameful'. But is it an attractive alternative to sentimentalism in general? Our new dispositionalist might point to exchanges like the following.

> Rex: "I can't believe he's still wearing that ratty old jacket. Mark is shamefully poor."
> Flavia: "Nonsense—he's poor because he's a graduate student. There's nothing shameful in that."
> Rex: "Well, I'd be ashamed to go around looking like that."

The speaker-relative dispositionalist says that Rex's original claim was nothing more than a statement about the tendency of poverty to elicit feelings of shame in him, in the case where he's the poor one. And perhaps this interpretation garners some support from the principle of charity, since

it treats Rex's report of his likely feelings as a demonstration of just the disposition he was claiming poverty had—that is, the disposition to make him feel ashamed.

The rational sentimentalist replies that Rex's second claim should be understood as a retreat from the primary and most obvious reading of the first. Having initially made an announcement about the fittingness of shame, Rex has backed up to something safer, but less interesting. (It's worth noting that this is a common phenomenon, and often the retreat is merely tactical. Rex may well still be convinced that shame is fitting, but he's not inclined to pursue the matter.) Not everyone finds this sentimentalist interpretation obviously preferable, however. At this point the dispute may seem a wash to some readers. But now the regulative role argument offers us something more to say. The sentimentalist will urge that if we took 'shameful' as shorthand for 'shame-inducing-for-me', then we'd need a different word with which to argue over and reflect on what to be ashamed of. But since, in other contexts, disagreement over what is 'shameful' so frequently seems precisely to be disagreement over what to be ashamed of, and since such disagreement would be so pointless on a speaker-relativized dispositionalism about 'shameful', why not insist that we already have such a word, and it's 'shameful'?

I suspect that the reason we use an apparently non-relative property term to express this concept is because the function of the concept is to allow us to pursue intersubjective emotional agreement by treating fitness-for-shame as a speaker-invariant feature of certain human characteristics. In other words, our mutual susceptibility to feelings of shame and our need for standards about when to feel it constitute good reasons for taking as central those uses of the term on which it serves the function of focusing thought and discussion about what to be ashamed of. And it is important to leave as a question of substance whether those standards should or should not recapitulate any particular person's (including one's own) present tendencies to be ashamed. If, as I think, these considerations lend support to sentimentalism, then we have begun to see how the kinds of considerations the regulative role argument invokes are relevant to choosing an interpretation of our actual evaluative terms.

The regulative role argument depends crucially on the thought that there are specific features of our emotional propensities that give us good reason to reflect upon and discuss standards for what to feel. Let me now try to say what those features are. One simple one is the connection between emotional experiences and motivation. States such as anger, envy, and shame, for instance, involve motivational tendencies: toward retaliation, competition, or concealment, respectively. Because it matters very much to each of us how we act, there's reason to think about what to be angry, envious or ashamed of. These facts generate an important role for intrapersonal criticism and reflection that an agent can undertake concerning the

appropriateness of these irruptive, motivating emotions. Furthermore, as Gibbard (1990) emphasizes, social life calls for coordination of action in very many human activities. And coordinated activity requires coordinated motivations. To avoid potentially expensive conflict, we do well to be able to agree on standards for when to feel emotions like anger and jealousy. And to avoid the disdain of others, and the further costs that brings, we must either live up to or convince them to alter their standards of shamefulness. These coordinative purposes supply a rationale for interpersonal criticism and reflection.

In addition to their direct motivational impact, emotions are ways of being bothered or pleased by things. Such pleasures and pains characteristically influence our other attitudes toward the objects of our emotions. Moreover, when we are in the grip of these valenced experiences, it is very natural to suppose that the situations on which we are focused matter in various distinctive ways. So emotions have a tendency to insinuate themselves into more richly conceptualized evaluative stances. Repeatedly finding oneself annoyed at a colleague's intrusions when one is working, for instance, may draw one toward views about appropriate office conduct that one would previously have thought stifling. In ways like this, emotional experiences influence the evaluative and regulative ideals that guide our planned behavior and our practical reasoning more generally. It is therefore vital to be able to reflect on particular such influences, to criticize or endorse them. A vocabulary in which to assess the fittingness of emotional responses supplies a valuable vehicle for such reflection.

These points bring us to another crucial feature of emotions which has been largely presupposed in what I've been saying already: namely, their idiosyncratic relationship to high-level cognitive processes like rational reflection. It has been noticed by various writers that emotions seem to be both perception-like in their independence from judgment and yet responsive to evidence and rational criticism in something like the way that judgments are.[7] In a slogan, emotions are independent from and yet responsive to reason. What has not been recognized is the role this fact plays in motivating sentimentalism.

Consider first the independence claim. Emotions are like perceptions in that they can arise independently of our considered convictions about the circumstances eliciting them, and they may even conflict with those convictions. We cannot simply decide to stop feeling them, nor can we always force them into line with our considered opinions. Furthermore, when they persist despite those opinions they induce us to question the opinions.[8] So even if I think beauty is only skin deep, and being smart or interesting or funny is what's important, I can be brought to think that one's appearance matters more than I previously acknowledged by finding myself ashamed of my flabby stomach at the beach. And even if I think envy, or jealousy, are contemptible emotions that I would be better off without, I can't be rid of them by wishing it were so. I do better to acknowledge the importance of the

concerns they embody, and seek ways of integrating those concerns with other important ideals. The apparent ineliminability and the autonomy from judgment of our emotions (i.e., their 'independence') indicate that we should expect our lives to continue to involve experiences that sometimes challenge our considered convictions—prompting us to acknowledge new sources of value, and perhaps even new categories of value. The fact that our emotions are independent in these ways is an important motivation for the regulative role argument. If change in emotional response reliably followed in the wake of change in judgment, there would be less need to turn our critical attention on our emotions. Because it does not, we need to think about whether our recalcitrant emotions are fitting responses.

On the other hand, it is equally important to note that emotional reactions *are* amenable to *some* measure of rational control or correction despite their independence—and this now is their 'responsiveness to reason.' For instance, someone more committed to the sufficiency of intellectual virtues might be able, by rehearsing his reasons for that commitment (and then choosing his companions carefully), to weaken or even eliminate feelings of shame at the beach. Similarly, cognitive behavioral therapy has had some marked successes in helping people overcome fear of flying, in part by persuading them of its comparative safety. If such changes were impossible, reflection on the appropriateness of shame or fear would be of merely theoretical interest. But since, in fact, reflection on how it makes sense to feel does exert some real influence on how we feel (especially over the long run), we have to decide how to use such control as we are able to exert. This is grist for the regulative role argument, since it motivates a role for a vocabulary in which to engage in and discuss such reflection about what to feel. It is puzzling how emotions can be both independent from and also responsive to reflection in the ways they seem to be. I will have a bit more to say about this later. For now it suffices to note that they are this way, and that this fact is of the first importance for the regulative role argument, and ultimately for sentimentalism.

All those seem to me good reasons for wanting some sort of vocabulary with which to think and talk about how to feel. But I still need to say something about why the nature of our emotional responses motivates a role for a vocabulary of terms expressing judgments of fittingness in particular. Why not just use ordinary exclamatory or imperatival vocabulary to express all-in judgments about what to feel? ('Boo for shame' or 'Feel angry.') This is an important question to the extent that the regulative role argument is supposed to favor rational sentimentalism in particular, as against other possible sentimentalist accounts.[9]

One answer comes from the phenomenology of emotional experience: it's in the nature of these experiences to present themselves as sensitivities to something outside them. And what they present themselves as sensitivities to is a fairly restricted feature of the situation: a socially significant personal inadequacy, or a threat to one's safety, for instance. A little introspection

makes it obvious, I think, that feelings of shame, fear, and so on just aren't about the advisability, or the moral permissibility, of feeling precisely that way. They are about a feature of the circumstance in virtue of which this is a fitting way to respond. In fearing the wild animal, I am struck by the things that make it fearsome (its size and ferocity, say), not by things that might make it wise or virtuous to be afraid. It may be neither wise nor virtuous to be afraid—if, for instance, I'm the only thing between the wolves and the children. But one wants a way to express the important sense in which fear is nonetheless appropriate to this situation: the sense that, whatever other reasons there may be for or against feeling it, the way this emotion evaluates the circumstances gets something right. Assessments of fittingness are attempts to make sense of or criticize our emotions using standards that speak to the distinctive concerns we take them to embody. It is therefore important to have a vocabulary that expresses such assessments, in particular, as a vehicle for rational interpretation of ourselves and one another.

A related point is also important. The extent to which various sorts of reasons for (and against) feeling some way are capable of influencing our actual reactions seems to vary depending upon whether they are reasons of fittingness or other sorts of reasons. Strategic reasons not to be afraid have almost no chance of calming our fears, whereas considerations about whether the circumstance is genuinely threatening do seem to have some influence.[10] So a vocabulary for expressing fittingness judgments is more likely to be an effective instrument in regulating our reactions than a vocabulary for expressing all-in endorsement. Furthermore, the considerations relevant to fittingness have a kind of contextual independence that other considerations about what to feel do not. This renders them more natural candidates for expression in an apparently property-ascribing discourse.[11]

The regulative role argument arises from the character of emotional experience, and from the fact that our emotions are concerned with aspects of life that have enduring significance to us. It is motivated in part by an acknowledgment that emotions seem to be or involve modes of evaluation that are distinct from, and sometimes in tension with, reflective evaluative thought. The sentimentalism I am advancing invites us to see regulative concepts, and the terms that express them, as a way of drawing some of these deep-seated (though sometimes unconscious, and sometimes disagreeable) human concerns into the ambit of language and reason—where they can be exposed, interpreted, considered, influenced, and partly, but not entirely, governed.

III. Disagreement and Essential Contestability

A different argument for sentimentalism arises from problems about evaluative disagreement, which contemporary sentimentalists have thought themselves in an especially good position to explain. Of course, philosophers

dispute how central disagreement is to our evaluative practices. But sentimentalists are typically impressed by the frequency and intractability of evaluative disputes. Several points should be emphasized in this connection. First, non-evaluative disputes can often be resolved by settling obviously empirical questions, or by stipulating a shared meaning for a contested term. But settling obviously empirical questions often leaves evaluative matters unresolved, and we are loath simply to stipulate what counts as, say, 'cheating' or 'deserved,' in order to resolve them. Furthermore, disputes over the application of evaluative concepts in particular cases often result from more general disagreements. The parties differ not simply over whether the concept applies here, but over what something has to be like in order to be the kind of thing to which the concept applies at all.

Take 'wrongness' for instance. People disagree not only over whether particular acts are wrong, but also about what features an act needs in order to be wrong. And philosophers seem no closer to agreement than the folk about how to settle such disputes. Thus a consequentialist and a deontologist famously can agree that torturing someone to find the location of the bomb would maximize the nonmoral goodness of the expected consequences, yet disagree over whether the act is wrong. Indeed, they may agree about all the nonmoral circumstances. Their disagreement over the case is founded in a more general disagreement about the character of wrongness. Similar remarks seem to apply, mutatis mutandis, in the case of beauty, shamefulness, funniness, and very many (perhaps all) evaluative properties. In each case, disagreement seems to be common not only over the extension of the property, but also over the proper conception of it. Furthermore, it is frequently not plausible to treat these disagreements as mere borderline cases of the sort that might engender difference of opinion for any number of nonevaluative concepts as well. Disputes over what acts are wrong (traits shameful, pictures beautiful, ...) arise even in cases that one of the parties regards as a central or paradigmatic application of the concept. And still the parties think that there remains a real disagreement; they do not suppose that they are simply talking past each other.

These phenomena are at the root of the suggestion that evaluative concepts are 'essentially contestable' (or, 'essentially contested').[12] But just what essential contestability amounts to has not been well explained. It must be stronger than the claim that the concepts are vague. No doubt evaluative concepts do admit of borderline cases, but that is not what's distinctively contestable about them. The idea seems rather to be that it is an essential feature of certain concepts that there is room for dispute over their application without linguistic impropriety, even in cases which one party to the dispute regards as clear or paradigmatic instances. Further, there is no guarantee that a dispute over an essentially contestable concept can be settled simply by appeal to the rules of application for the concept together with the nonevaluative facts of the case. Yet philosophers who think that

evaluative concepts are like this characteristically insist that this is not yet to say that there is no possibility of error in application, nor even that both parties to such a dispute are rationally blameless. It is to say that if either party is guilty of error, that error must be located somewhere other than in a failure of conceptual competence or nonevaluative knowledge (often the error is attributed to a failure of *sensibility*). So understood, essential contestability also seems to be one of the ur-thoughts that attracts some philosophers to noncognitivist treatments of evaluative discourse, though many of its proponents reject noncognitivism.

I find the claim that evaluative concepts are essentially contestable plausible, but I cannot offer further support for it here. Instead, I want to note that this thesis sits uncomfortably beside a second intuitively attractive thought about value judgments: namely, that predicates like 'wrong' can be given a univocal interpretation across very different patterns of application and different theories of wrongness (i.e., that parties to the sort of dispute imagined above really are talking to, rather than past, each other).The univocity of evaluative predicates can seem to be threatened by the very existence of the sort of evaluative disputes mentioned above. What justifies us in thinking that parties with such different views about what things are wrong and what makes them be wrong are talking about the same thing when they apparently disagree over whether an action is 'wrong'?[13] Why not think instead that their use of the same word disguises a difference in what they are talking about—a difference manifested by their inability to agree on what further sorts of considerations would settle the dispute?[14]

Here's where sentimentalism might offer some help.[15] Perhaps what secures univocity in a radical dispute, despite the essential contestability of the concept, is a common sentiment that somehow supplies a shared subject matter for the discussion. As David Wiggins (1987, p. 198) puts it "we can fix on a response . . . and then argue about what the marks are of the property that the response itself is made for. And without serious detriment to the univocity of the predicate, it can now become essentially contestable what a thing has to be like for there to be any reason to accord that particular appellation to it, and correspondingly contestable what the extension is of the predicate." The idea is that a shared response, or sentiment, somehow moors us in a common subject matter, making it possible for us to disagree substantively about what a thing has to be like in order to be such that we should feel *this* sentiment toward it.[16] Thus, if the sentimentalist is right, it is because our evaluative concepts have a special tie to shared human sentiments that we are able to engage meaningfully in debates over their application. And the point of these debates essentially involves the regulation of a particular kind of emotional reactions to the world. A shared sentiment supplies a shared element in the intensions of our evaluative thoughts.

Notice that this proposal imposes a crucial constraint on a sentimentalist account of emotions. If univocity for the discourse is to be secured despite the essential contestability of its concepts by shared sentiments whose appropriateness the disputes are about, then sentimentalism owes an account of these sentiments that makes it clear that, at least in some cases, parties to these disputes are parties to the same sentiments. Discharging this debt would require a general account of what these sentiments are. But minimally, we must be given some reason to have greater confidence in the claim that parties to a dispute over what's Φ share the same emotion F than we had already in the claim that they mean the same thing by 'Φ' before we took this excursion through their emotional repertoires. Absent any such reason, we are no better off than we began with respect to the problem of univocity.

Here McDowell and Wiggins seem to be in arrears. Wiggins (1987, p. 195) insists, and McDowell (1997b, p. 219) seems to agree, that the responses to which sentimentalism appeals cannot be understood as conceptually 'prior to' the evaluative concepts or properties these responses are invoked to explain. Property and response are equal partners. So, when they come to the question of what these sentiments are, each of these authors suggests that any adequate answer to it must appeal to the properties to which the sentiments are responses. There is no saying what amusement is without appeal to the funny, and no saying what shame is without appeal to the shameful, etc. This raises worries about circularity, of course. But even if one grants that circular elucidations are sometimes informative, still, to identify the sentiments by appeal to the properties to which they are appropriate responses would surrender any promised advantages for securing univocity alongside essential contestability. If the parties to a given dispute have different views about a property's extension and different views about what other features make something have the property, then the initial worry about univocity just is a worry about whether they are speaking about the same property. So if the sentiment that each is feeling can only be individuated by appeal to the property to which she is responding, then the claim that they are talking about the appropriateness of the same shared sentiment is no more secure than the claim that they are talking about the same property (or deploying the same concept) was at the outset. I conclude from this that sentimentalism only makes headway on the problem of univocity if the sentiments it invokes can be identified independently of the evaluative properties they putatively respond to or fit.

IV. Sentimentalism and the Emotions

The arguments offered above impose some significant constraints on the states to which a sentimentalist account of evaluative concepts as regulative concepts can appeal. First, the sentiments must be states that

are widely enough shared to supply a common subject matter among discussants with competing views of their fittingness. The more widely they are shared, the better suited they will be to ground evaluative discussions across differences in sensibility. Second, to make headway against the problem addressed in the previous section, it is important that the basis for attributing sameness of sentiment does not presuppose sameness of evaluative properties or concepts. Third, the sentiments must be sufficiently independent from high-level evaluative judgment to be capable of conflicting with it—this motivates the need for reflection about their fittingness. But, fourth, they must also be somewhat responsive to reflection on their fittingness, or such reflection would be otiose, like reflecting on whether to find sugar sweet. Finally, the regulative role argument will be more persuasive to the extent that human beings are stuck with the sentiments in question, since such ineliminability ensures our ongoing interest in reflecting upon and attempting to guide our emotions.

Any sentimentalism motivated along the lines I have suggested here requires an account of the emotions that meets the constraints above. In the space remaining I briefly consider some consequences of these constraints. I suggest that familiar philosophical attempts to define emotions by appeal to their constituent propositional attitudes generate a class of emotions that does not meet the above constraints well. Whereas, states that I will call 'natural emotions' do appear to meet those constraints remarkably well. This suggests that the question of which evaluative concepts a sentimentalist theory might be true of depends in part on which human emotions turn out to be natural emotions. If I'm right about each of these points, then the prospects for and the final shape of a sentimentalist theory depend on the outcome of empirical investigations of emotions.

The dominant so-called 'cognitivist' tradition in the philosophy of emotion has held that emotions are constellations of propositional attitudes, perhaps conjoined with feelings or physiological components. On this view, emotions are to be individuated by differences between the propositional attitudes (typically, beliefs) that are their essential constituents. I think this tradition is misguided in various ways, not all of which are relevant here. But I would grant that there are some states that are helpfully explicated at least in part by appeal to the cognitivist model. I call emotions that are identified by appeal to constituent propositional attitudes 'cognitively sharpened emotions.'[17] The relevant issue at the moment is this: I suspect that cognitively sharpened emotions will not play the role required of the sentiments in the arguments I've been considering.

There are two reasons for this suspicion. First is the role of emotions in securing univocity. To define an emotion by appeal to its propositional content is to require that in order to count as feeling this emotion one must have the relevant content in mind, in some sense. So two parties will only count as feeling the same cognitively sharpened emotion if they each

have attitudes involving that same content. But this is precisely to eschew using the sentiments themselves to secure that shared content. So the sentimentalist argument from univocity seems to demand a different conception of emotions than cognitively sharpened emotions.

Secondly, there are grounds for doubting that cognitively sharpened emotions have the independence from higher-level cognition I appealed to in defending the regulative role argument. Recall that it was in part the fact that emotions arise unbidden, sometimes challenging our evaluative judgments, that generated a need for a vocabulary with which to discuss and reflect upon their fittingness. But there is no reason to expect that this would occur with cognitively sharpened emotions and some reason for doubting it. If the emotions were properly understood as requiring the deployment of the conceptual, linguistic resources characteristic of propositional attitudes, there would be no reason to expect emotional experiences to provide a source of evaluations that diverged from and demanded reconciliation with higher cognition. In which case there would be nothing to prompt the reflections that motivate the regulative role argument. I believe that those emotions that arise unbidden and sometimes seem to conflict with evaluative judgment typically do so because they are not themselves exercises of our capacities for thinking in language, but products of fundamentally different, and evolutionarily more ancient, evaluative systems.

Now consider, in contrast, the natural emotions. Natural emotions are heritable suites of cognitive, affective, motivational and behavioral changes that are part of the normal human repertoire in every culture because of our shared evolutionary history. These syndromes are products of relatively discrete special-purpose mechanisms that are sensitive to and focused upon various important aspects of human life. Paul Ekman and his colleagues have been at the forefront of research on such states, and have found distinctive physiological profiles that differentiate a number of them.[18] At the moment that number is small—six or seven—but the research program is young and its methods to date give some reason to suspect substantial undercounting. The heritability and ubiquity of these complex and coordinated responses lend some credence to the idea that they have an evolutionary history, though of course anything we say at this stage about their adaptive functions will be speculative. Ekman (1980), Robert Zajonc (1980), and others treat the emotions as 'automatic appraisal mechanisms' or 'affective information-processing systems.' It is helpful to think of them as modular, in Jerry Fodor's (1983) sense of the term. They are, to some extent, informationally encapsulated in the kinds of subject matter they accept as input, cognitively impenetrable in that they are often unresponsive to conclusions generated in other parts of the cognitive system, and mandatory in that the reactions are typically not amenable to direct voluntary control. Obviously, the claim that any given emotion is a natural emotion in

this sense is a speculative empirical claim. But recent research seems to support some such claims.[19]

Consider fear, an emotion that has commonly been held to be a natural emotion. What might explain the integration of the distinctive affective, motivational, physiological and cognitive changes that are involved in fear? At one level, the answer is a cluster of physical mechanisms that, in the case of fear, are relatively well understood. (LeDoux, 1996.) For present purposes the details of these mechanisms don't matter, what matters is that there are some mechanisms or other that consistently cause the integrated co-occurrence of these symptoms in normal human beings. At a second level of explanation, the answer is that a history of natural selection has supplied us with adaptations that dispose us to react in certain ways to situations of the sort that characteristically cause all these symptoms, because the co-occurrence of such reactions to such situations tended to increase the genetic representation of the creatures that had them. But what sorts of situations are these?

Differences between individuals and cultures may generate radically different response profiles, but reflection on the adaptive role of these responses reveals similarities amidst the differences. A contemporary American experiences fear when her car doesn't respond to pressure on the brake pedal, whereas ancient Greeks experienced fear in response to certain patterns of entrails. But these reactive differences are compatible with a deeper similarity: they are each products of a system that reacts differentially to what an observer might call 'perceived dangers'. But we do not need to be in a position to attribute shared concepts in order to attribute a shared emotion type. As our biological understanding of these syndromes advances, we acquire grounds for attributing them to parties with very different sensibilities.

Talk of modularity might lead us to expect that fear would be elicited only by specific visual cues, of advancing predators, for instance. But it is clear that this is false. Higher cognitive processes such as inference can initiate fear, as when one works out that there is not enough fuel in the tank to take the plane to the nearest landing strip. Such processes can also fail to permeate the fear system though: no amount of reflection on the dangers of smoking seems able to make those little white sticks frightening (as even those who manage to quit can attest). How and why some reflections and not others are able to penetrate emotional systems are puzzling questions, but not ones that I expect philosophical reflection to answer for us.

Now how does all this help with sentimentalism? The hypothesis that fear is a natural emotion gives us reasons for confidence that adopting an idiosyncratic view about what kinds of dangers to care about, for instance, would not make one no longer susceptible to the familiar human emotion of fear. Even someone who denies that the prospect of serious physical injury is fearsome can be understood as disagreeing with us conventional cowards

over the appropriateness of fear toward it. Despite the differences between the inputs that make us afraid and the ones that make him afraid, so long as fear is a natural emotion, we can treat him as a party to that emotion that we share, and we can treat his capacity for that emotion as his anchor in our discourse over what's fearsome.

Natural emotions are "pancultural" as Paul Griffiths (1997) puts it. Though some sociopaths and persons with brain injuries may lack the capacity for these responses, normal human beings in every culture are subject to them. This offers the prospect of meaningful evaluative disputes across striking differences in cultural practice. The modularity of natural emotions is equally important. It means that they are to some degree independent of high-level evaluative reflection. This independence underwrites the need mentioned earlier to reflect upon their fittingness, which in turn helps to explain the evaluative vocabulary that has arisen to meet that need. But the autonomy of emotional evaluation systems from judgment and inference is only partial, as we've seen. Fear, shame, anger, envy, etc., are not as impenetrable as that paradigm of modularity, the visual system. They are sometimes responsive to reason. Thus, despite their independence from higher cognitive processing, there remains a point to reflection and discussion on their fittingness, and, again, to the vocabulary that subserves that discourse. Finally, it may be that natural emotions are ineliminable for practical purposes: an inevitable fact of life under normal conditions of human development. This ineliminability would lend further force to the regulative role argument. It is partly because we are stuck with these recurring forms of reaction that we need a vocabulary with which to reflect upon and converse about what circumstances are such as to make the reactions appropriate.

Conclusion

The regulative role argument and the argument from essential contestability seem to me to be significant and distinctive motivations for a sentimentalist approach to certain evaluative concepts. But they depend upon some rather restrictive assumptions about the nature of the sentiments. I have tried to make those assumptions explicit, and to offer some reasons for thinking that they are plausible with respect to natural emotions, at least. These arguments seem to me to lend some needed support to sentimentalist accounts of value. The arguments also suggest, though, that for a sentimentalist account of a given concept Φ to be plausible, the sentimentalist must identify a determinate sentiment F that satisfies the desiderata that emerged above.

Sentimentalism has traditionally been promulgated and assessed as an account of evaluative concepts quite generally. But in view of these

considerations, it may be wiser to adopt the doctrine piecemeal with respect to just those specific evaluative concepts that are paired with the right sorts of emotional responses. Which brings us to the questions of just which concepts that might be true of, and whether, as Allan Gibbard (1990) has argued, there is reason to think it will be true of moral concepts in particular. The forgoing considerations suggest that this is in part an empirical question. It depends upon whether there are emotions that meet the restrictions advanced here and are such that reflection on right and wrong can plausibly be interpreted as reflection on their fittingness. The present state of empirical research on emotions leaves this a wide open question. If the forgoing lines of thought are on the right track, then answering this empirical question has some very important consequences for moral theory.[20]

Notes

1. I will alternate between talk of concepts and of judgments applying those concepts, as convenient, in cases where I think differences between these are irrelevant to the issue at hand. By evaluative 'judgment' I mean the mental state of reaching an evaluative verdict—whether that state is understood as a belief or some noncognitive attitude, and whether or not it is publicly expressed. Evaluative concepts are constituents of such verdicts.
2. This question is discussed in more detail, and further candidate answers are vetted, in D'Arms and Jacobson (forthcoming).
3. D'Arms and Jacobson (2000b). For similar suggestions see Rabinowicz and Rønnow-Rasmussen (2004).
4. Perhaps a sophisticated dispositionalist could avoid this consequence as well, by judicious specification of the conditions under which people are disposed to be ashamed of the shameful, or of the people whose dispositions are determinative. This is worth a try, but I doubt that any substantial specification of these conditions or responders can be found which will allow the analysis to deliver plausible verdicts in a sufficiently wide range of cases.
5. See Brandt (1978), Lewis (1989), Railton (1989), Gibbard (1990).
6. This possibility is explored in Railton (1989).
7. See, for instance, de Sousa (1987) and Greenspan (1988).
8. The role of emotional experience in evaluative revision has recently been explored in interesting ways by Bennett Helm (2001).
9. As presented, the argument is offered in support of rational sentimentalism in particular. But it could easily be adjusted, by making premise three more ecumenical, to support any of the second-order sentimentalist positions mentioned earlier. My strategy throughout this paper is to present the arguments in a way maximally favorable to my preferred version of sentimentalism, but I believe something like the arguments I offer here would be embraced by most sentimentalists.
10. See D'Arms and Jacobson (2000a).
11. One might hope to explain the inefficacy of strategic reasons by thinking about the evolutionary function of emotions. Franks (1988), offers such an explanation

with respect to anger. If we could get ourselves to stop feeling anger when it wasn't in our interests to feel, it couldn't play the role it plays in deterrence. Analogous functional arguments suggest themselves for some other emotions, such as jealousy, shame, guilt, and fear, though these arguments depend upon many assumptions it would be difficult to establish.

12. See Gallie (1956), Hurley (1989), Wiggins (1987). Sabina Lovibond (1983) also seems to embrace the suggestion, without using the term.
13. David Merli (MS) develops the univocity problem incisively against moral realism.
14. It is tempting to answer that we can easily identify a common subject matter for disputes over wrongness by supposing that what is at issue between the parties is what one ought to do. But if the ought in question is a rational ought, then there is no guarantee that this is what they disagree about, since the rational force of moral demands is itself a subject of apparently substantive dispute. Whereas if the ought in question is a moral ought, then those who are worried about the univocity of 'wrong' are going to have the very same worry about the univocity of this 'ought'.
15. There are, of course, non-sentimentalist strategies for attempting to secure univocity in the face of widespread and intractable disagreement, which face various difficulties I can't address here. Most strive to explain away, rather than accommodate, the appearance of essential contestability.
16. Cf. Hare (1952, chapter 7) on the primacy of the commending meaning of 'good'.
17. The distinction between natural emotions and cognitive sharpenings is explained and defended in more detail, and the cognitivist tradition is criticized, in D'Arms and Jacobson (2003).
18. See Ekman (1980, 1993) and Ekman and Friesen (1971).
19. For a helpful review of much relevant literature on basic emotions, see Griffiths (1997).
20. I'm grateful to audiences at Ohio State University, Rutgers University, University of Wisconsin, University of North Carolina, and a Symposium on Empirical Approaches to Ethics at the Pacific APA, for useful discussion of ancestors of this paper; and to John Doris and Sigrun Svavarsdottir for helpful comments on an earlier draft. Above all I am indebted to Daniel Jacobson for discussions in which many of the ideas here were developed.

References

Anderson, Elizabeth (1993). *Value in Ethics and Economics* (Cambridge, MA: Harvard University Press).

Blackburn, Simon (1993). *Essays in Quasi-Realism* (New York: Oxford University Press).

Brandt, Richard (1979). *A Theory of the Good and the Right* (Oxford: Clarendon Press).

D'Arms, Justin and Daniel Jacobson (2000a). "The Moralistic Fallacy," *Philosophy and Phenomenological Research* LXI,1: 65–90.

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

D'Arms, Justin and Daniel Jacobson (2003). "The Significance of Recalcitrant Emotions (Or Anti-QuasiJudgmentalism)" in *Philosophy and the Emotions*, A. Hatzimoysis, ed., (Cambridge: Cambridge University Press).

D'Arms, Justin and Daniel Jacobson (forthcoming). "Sensibility Theory and Projectivism" in *Oxford Handbook of Ethical Theory*, David Copp, ed. (New York: Oxford University Press).
de Sousa, Ronald (1987). *The Rationality of Emotion* (Cambridge, MA: MIT Press).
Ekman, Paul (1980). *The Face of Man* (New York: Garland).
Ekman, Paul (1993). "Facial Expression and Emotion," *American Psychologist* 48(4): 384–392.
Ekman, Paul and Friesen, Wallace (1971). "Constants Across Cultures in the Face and Emotion," *Journal of Personality and Social Psychology* 17(2): 124–129.
Fodor, Jerry A. (1983). *The Modularity of Mind* (Cambridge, MA: MIT Press).
Frank, Robert (1988). *Passions Within Reason: The Strategic Role of the Emotions* (New York: W.W. Norton and Co.).
Gallie, W.B. (1956). "Essentially Contested Concepts," *Proceedings of the Aristotelian Society* 56: 167–198.
Gibbard, Allan (1990). *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press).
Greenspan, Patricia (1988). *Emotions and Reason: An Inquiry into Emotional Justification* (London: Routledge & Kegan Paul).
Griffiths, Paul (1997). *What Emotions Really Are* (Chicago: University of Chicago Press).
Hare, R.M. (1952). *The Language of Morals* (New York: Oxford University Press).
Helm, Bennett (2001). *Emotional Reason: Deliberation, Motivation, and the Nature of Value* (Cambridge: Cambridge University Press).
Hurley, Susan (1989). *Natural Reasons* (New York: Oxford University Press).
LeDoux, Joseph (1996). *The Emotional Brain* (New York: Simon and Schuster).
Lewis, David (1989). "Dispositional Theories of Value," *Proceedings of the Aristotelian Society* Supp. LXIII: 113–137.
Lovibond, Sabina (1983). *Realism and Imagination in Ethics* (Minneapolis: University of Minnesota Press).
McDowell, John (1997a). "Values and Secondary Qualities" in Darwall, Gibbard, and Railton, eds. *Moral Discourse and Practice* (Oxford: Oxford University Press).
McDowell, John (1997b)."Projection and Truth in Ethics" in Darwall et al, 1997.
Merli, David (MS). "Moral Realism's Semantic Problem."
Mulligan, Kevin (1998). "From Appropriate Emotions to Values," *The Monist* 91,1: 161–188.
Rabinowicz, Wlodek and Toni Rønnow-Rasmussen (2004). "The Strike of the Demon: On Fitting Pro-attitudes and Value," *Ethics* 114: 391–423.
Railton, Peter (1989). "Naturalism and Prescriptivity." *Social Philosophy and Policy* 7: 151–74.
Wiggins, David (1987). "A Sensible Subjectivism?" in his *Needs, Values, Truth* (Cambridge, MA: Blackwell).
Zajonc, Robert (1980). "Feeling and Thinking: Preferences Need No Inferences," *American Psychologist* 35: 151–175.