Moral Feelings and Moral Concepts

Allan Gibbard

In my 1990 book *Wise Choices, Apt Feelings*, I set out to analyze the narrowly moral concepts morally wrong and morally reprehensible.\(^1\) Wrongness pertains to acts, to what a person does, whereas being blameworthy or reprehensible pertains to the act along with motives and the agent’s state of mind.\(^2\) A man in the pangs of grief, imagine, speaks woundingy to a friend who is trying to bring comfort. The act might well be wrong and yet, in light of his wrought-up state, he might not be acting reprehensibly. The concept reprehensible I treated as the more basic of the two, defining the word ‘wrong’ in terms of an action’s being reprehensible.

A chief part of my interest in these narrowly moral concepts was, of course, to try to understand what is at issue in questions of right and wrong and in questions of blameworthiness. Another motive, though, was to discern what’s at issue in blanket attacks on ‘morality’, as with Nietzsche and Williams.\(^3\) Can we understand morality, in a narrow sense, as something we could be without, and that other cultures perhaps do lack, for better or worse?

Narrowly moral questions, I proposed, concern how to feel about things that people do or might do. For an act to be morally wrong isn’t just for it to be inadvisable or against the demands of reason. Passing up an enjoyable

---

\(^1\) I use small caps to denote concepts. This is a variant of the convention that Horwich adopts (1998).

\(^2\) Ross (1930) distinguishes ‘acts’, apart from motives, from ‘actions’, which are acts with their motives. Brandt (1959) discusses blameworthiness as distinct from wrongness.

\(^3\) Nietzsche (1887); Williams (1985).
I regard as the basic normative concept, the conceptual atom that renders molecular concepts like reprehensible normative. Much in the account of narrowly moral concepts that I have sketched must hinge on how this concept warrant is explained. But mostly, in the rest of this paper, I'll be pursuing questions that arise independently of how we gloss warrant. We could be non-naturalists about warrant, as Ewing was, or dispositionalists about warrant, as we might read philosophers in the ideal observer tradition, such as Firth, Brann, and Michael Smith, as being. I'll briefly sketch my own account of the concept warrant, though, for two reasons. First, it bears on the very point of moral evaluation: narrowly moral questions, if I'm right, are questions of how to feel about things people do or might do. Second, I'll later be asking whether the right view to take of morality is “realist”. I need, then, to say in what ways my own view is a form of realism and in what ways it isn't. Joined with a realist account of the concept of warrant, the analysis of reprehensible I have offered would come out as realist, whereas joined with an irrealist account of warrant, this analysis of reprehensible would be irrealist—for it would have an irrealist component. As for how to classify my own analysis, the question is tricky; my analysis is what Blackburn calls “quasi-realist”.

Acts, beliefs, feelings—each of these sorts of things, it seems, can be warranted or not. We speak of the thing to do, the thing to believe, or the thing to feel. We can ask what it makes sense to do, what it makes sense to believe, and how it makes sense to feel about a person or an action. My account of what's at stake in such questions is this; begin with action: the basic normative question is what to do, and to come to a conclusion on this is to form a plan. On a trip, say, I ask whether it makes sense to explore the city this afternoon or to rest first. I come to an answer by forming a plan. The thing to do, I might conclude, is to rest first; to come to this conclusion is to plan, for this afternoon, first to rest.

The same goes for feelings, I argue. The question of how to feel about something is, in effect, a planning question. I ask myself, say, how to feel about lies told to criminal suspects during questioning. As I work toward an answer, I'm in effect forming a plan for how to feel in my circumstance. I can plan what to do, and in a way I can likewise plan what to believe given certain evidence: planning an experiment, statisticians stress, crucially includes plans for how to analyze the outcome, contingency plans for what to believe should the experimental data turn out in various ways it may turn out. The same goes, I claim, for feelings.

4 Ewing (1939); Firth (1952); Brandt (1959); Smith (1994).
5 Blackburn (1993); (1998). I adopt this term in Thinking How to Live.
This analysis is in essence the one I gave in *Wise Choices* in 1990, and in my new book *Thinking How to Live* (2003), it is developed this way explicitly. The new book takes further a set of claims I made in the previous one. A planning concept of the kind I analyze would, it argues, act in ways that might have been thought diagnostic of normative realism. We are guaranteed, for instance, that there’s a non-normative property that constitutes being the thing to do. The theory starts out, then, with somewhat the same kinds of materials as does Ayer’s emotivism—a paradigm, we might have though, of anti-realism in moral theory. My theory differs from Ayer’s in important ways, to be sure, but the core of Ayer’s theory could. I claim, be put in a form that has exactly the same implications for questions of moral realism as does my own theory. (Indeed that, in effect, is what Simon Blackburn does with his own form of “quasi-realism”.) I don’t end up denying that moral claims are true or false. I do end up claiming that there are moral properties, and that these are natural properties in a liberal sense of “natural”. Indeed these properties could, for all my metamorphic theory tells us, be properties subject to empirical investigation by psychologists and social scientists. In crucial ways, then, the theory ends up very much like the non-naturalistic moral realism of G. E. Moore—and it also endorses some of the central claims of naturalistic moral realists like Richard Boyd.

All this, however, is by way of a priori investigation into concepts. The new book isn’t much about us, the flesh and blood organisms whose concepts these might be. In this present paper, I’ll return to psychology, to the bearing of empirical investigations into our nature on metamorphic questions, on questions of what our normative concepts might be.

Wrong and Guilt: Empirical Concerns

I’ll begin with an objection to my theory by Sean Nichols, in his book *Sentimental Rules* (2003). My book *Wise Choices*, unlike the new book, was highly concerned with the actual human psyche and with narrowly moral concepts. Among other things, as I indicated, it analyzes the concept of wrong in terms of reprehension and guilt. This can’t be right, Nichols argues, for young children acquire the narrowly moral concept of wrong long before they have the concept of guilt. Wrong they understand by the age of 4, but guilt not until they are 6. Their moral concepts at the start, then, can’t be the ones I depict. That, he argues, discredits my analysis too as it applies to adults. If, after all, our concept wrong isn’t one that we share with 4-year-old children, then we aren’t agreeing or disagreeing with them when we wield our own concept.

Similar grounds for worry about my analysis stem from anthropology. When I was writing *Wise Choices, Apt Feelings*, I was frustrated not to find much by way of reports on what people actually say to each other when they criticize, gossip, and the like. I likewise couldn’t find much on the moral emotions and the like that figure in people’s responses to each other. The situation is somewhat better by now, with some anthropologists asking very much the kinds of questions I want them to be asking. A lesson that may be emerging is this: On the one hand, moral outrage or something close to it is a human universal. On the other hand, guilt, as European-rooted cultures know it, may be culturally special. Again, we can ask, if some cultures lack guilt or lack a concept of guilt, are these people just using different concepts from ours? Does that mean that we can’t straightforwardly agree or disagree with them when they make outrage-laden judgments?

Talk of guilt and its absence cries out for some explanation of what we should take guilt to be. How does guilt, as we know it and hate it,
differ from other kinds of anxiety, such as fear and shame in their various forms. Guilt-absent cultures are by no means free of anxiety; the point is that anxiety at having done something wrong may, in those societies, lack certain special earmarks of guilt. Writers take a range of positions on what these earmarks are, but here is my own rough view of the matter. Guilt normally ‘meshes’ with the resentment of others, in that it tends to arise when one would resent others if positions were reversed and one knew what they had done. The feeling doesn’t require that others know what one has done—or, if they do know, that they do in fact resent it or may resent it. It doesn’t, moreover, respond just to the bare fact that others do resent one. Rather, it is governed by the same standards as govern outrage or resentment, though from the standpoint of the agent rather than the one affected. In roughly the ways that resentment requires thinking an act voluntary, so does guilt. Finally and crucially, guilt is characterized by its tendencies toward actions and displays that work toward reconciliation. One tends to display one’s pain to one’s victim and to bystanders, thus neutralizing any advantage one has gained at the victim’s expense. Feeling guilty also prompts one, if possible, to find a way to “make it up” to the person wronged.

Guilt can be over lack of due care, but the paradigm is guilt over a deliberate action that one now regrets. In the paradigm case, guilt thus requires a “change of heart”, and to display pangs of guilt to others is to display this change of heart. Take a child who pushes another child off a swing so that he can get the swing himself. He won’t feel guilty for what he did unless he now regrets it, preferring to have left the child alone and forgone the swing. Even guilt over an accident may display a change of heart on how much care to take to avoid the harm. We can understand guilt, then, only if we understand changes of heart, coming to respect the interests of others in a way that one previously didn’t. (Perhaps this is what the 5-year-old doesn’t comprehend and the 6-year-old does. We can ask, is this scenario absent from the lore of some cultures?)

9 My views are closely related to those of Baumeister et al. (1994) as well as to Fessler and Haley (2003).

10 Neither resentment nor guilt invariably require thinking an act voluntary, but the exceptions correspond for the two emotions. “Survivor guilt” notoriously needn’t stem from one’s voluntary actions. Yet it seems aptly named: the “flavor” is that of guilt, it focuses on a kind of alienation from one’s fellows, and it copes with a fact that may elicit their resentment. Guilt over actions of one’s group also doesn’t seem to require any act of one’s own, and so not, in particular, a voluntary act of one’s own. But that goes for resentment too, as in ethnic rampages where members of a group are attacked and often killed over acts of other members. Guilt and resentment may well not be warranted in those circumstances or may be misdirected, but these do seem to be recurring emotional patterns.

Guilt is closely tied to anxiety over social exclusion, over alienating those who are important to one. But social exclusion will be disastrous anywhere, and so anxiety over alienating others must note be a human universal. Guilt focuses not directly on the social danger, but on the ways one has come to warrant the resentment of others. Not that guilt isn’t highly affected by the social perils of being resented, but with guilt one’s emotional focus isn’t on these dangers but on one’s misdeed.

This emotional style of dealing with one’s transgressions may not be found everywhere, and if it isn’t, then we can ask what it is about cultures rooted in Europe that inculcates such a style. One evident answer is the practice of apology, often forced on the young child and found painful and humiliating. How, then, absent the pattern I have described, do people deal with their own transgressions and the people they wrong? We need to learn more about this from ethnographers; it strikes me as an important question for understanding what’s universal in morality and what’s local to certain kinds of cultures.

If a culture lacks guilt, Nichols can say, then it must lack narrowly moral concepts as I analyze them. Now in a way, of course, the conclusion that cultures may lack narrowly moral concepts might fit in nicely with the project of identifying morality in a narrow sense, of identifying a target for the critiques of Nietzsche and Williams. Both these philosophers were concerned with guilt (along with a hyperbolic sense of free will and responsibility that won’t be my concern here). If parts of humanity manage without guilt, so much the better for the prospects of weaning us ourselves from guilt. But at the very least, for people with outrage but not guilt in their emotional repertoires, we need to ask what ethical issues we still can discuss with them, and what concepts we can draw on to formulate common issues.

Outrage without Guilt

Nichols, in his critique, on two sets of findings in developmental moral psychology. First, as Turiel and others have found, quite young children make a conventional/non-conventional distinction. Ask a child, “Suppose the teacher said that hitting was allowed in the classroom. Would it be all right to hit people?” Children by the age of 4 answer no. Ask

11 Baumgartner et al. (1994: 246) speak of guilt as having “two sources: empathic arousal and anxiety over social exclusion”. I perhaps should have included empathic arousal in my account, but I don’t have any special reason to think empathy absent in guilt-absent cultures.

12 Turiel (1983); Turiel et al. (1987).
Perhaps that's all right, and just shows the peculiarity of our own institution of morality, for better or for worse. If I got our own narrowly moral concepts right, it seems to follow, then we can't discuss questions of narrowly moral right and wrong either with young children or with many exotic adults.

Still, there are closely related issues that we could discuss with them, for all these findings show. It hasn't been shown, after all, that either of these populations lacks a concept of outrage—and indeed it would be surprising if they did. We can share with a wide range of people, then, a set of pared-down narrowly moral concepts. I call these "near-moral" concepts. An action is outrageous, in this near-moral sense, just in case outrage over it is warranted on the part of others. Suppose, then, as may well be, that the concept of outrage is a human universal, and that the concept emerges fairly early in a child's development. Then, for all we have seen, the quasi-moral concept of being outrageous may be available to all adults and to all children at a fairly early age. All they need is a concept of reprehension or outrage and a concept of warrant, and they have the ingredients for a near-moral concept of being reprehensible.

For us, the quasi-moral concept of being outrageous might be exactly equivalent to the narrowly moral concept, in the following sense. Here is something that we may perhaps accept as truistic, as not in need of debate and discussion: guilt over something one has done is warranted just in case outrage over it is warranted on the part of impartial observers. If I relax with some light reading instead of grading a set of overdue papers, I'm warranted in feeling guilty over what I have done just in case an impartial spectator would be warranted in feeling outraged over it. Clearly this isn't something that a young child could accept, or an adult from a guilt-absent culture. But perhaps it is something that we in a guilt culture do take for granted.

Now if that is so, then on our own view, an action is outrageous in the narrowly moral sense just in case it is outrageous in the near-moral sense. We think that guilt is warranted just when outrage is—that guilt is warranted on the part of an agent just when outrage is warranted from impartial others. We are committed, then, to this claim: guilt and outrage are warranted in just those cases where outrage is warranted. In other words, an action is reprehensible in the narrowly moral sense just in case it is reprehensible in the quasi-moral sense.

Naive Concepts: Wrong and Red

For anything the guilt critique shows, young children might think in near-moral terms in the way I have pictured. We haven't seen that they

---

lack the psychological concept of outrage or reprehension, and we haven’t seen that they lack the normative concept of warrant. The outrage-warrant approach is to build near-moral concepts out of these two elements. Still, it does seem implausible that the goings-on in young children’s heads, when they think in near-moral terms, are put together in the complex way the outrage-warrant analysis depicts. Something more direct may be going on. Indeed, the same may well be going on with most adults—even including us philosophers much of the time. I’ll sketch my best guess for what this something might be, and then spend the rest of the paper on the upshot if this guess is right.

What do the thinking of young children, of guilt-absent adults, and of the rest of us much of the time have in common? Start again with young children. They, like the rest of us, have emotional reactions to things like hitting. (With adults, indeed, we now know which emotional centers of the brain are active when one thinks, say, of pushing a man in front of a moving trolley.) Emotions, on the story that neurophysiologist Antonio Damasio tells, involve, among other things, perceptions of visceral and other bodily goings-on set off by certain kinds of thoughts. (This part of the story is reminiscent of Hume, and psychologist Jerome Kagan’s dustjacket praise on Damasio’s first book read “Hume must be smiling.”) The thought of one child hitting another, then, has a certain emotional “flavor”, as we might put it; the flavor is a matter of what it’s like to experience those bodily sensations and “bundle” them, as it were, into one’s whole conception of what goes on.

As the word ‘flavor’ suggests, all this is, in a way, like experiencing a taste of ice cream that one can recognize—though the path from the hitting to the emotional flavor may be more circuitous, presumably, than that from the ice cream to the taste. Instead of talking about flavor literally, though, let me consider another analogy, color vision. Certain patterns of electromagnetic radiation lead to an experience of red. And on seeing a thing as red, we naively attribute a property to the object seen; we take it that the thing is red. How exactly all this should be put is of course a perennial philosophical issue, which I won’t get much further into. My point here is that much of what can be said about seeing something as red and so thinking it red can be said in the moral case. Being outraged by an act is to thinking it wrong as seeing a thing as red is to thinking it red. That is the parallel I want to explore; it is familiar from Hume, Blackburn, ideal observer theorists, and others.

With color, just as with morals, as we all know, it is hotly debated whether the right story is a form of color realism, color irrealism, or error theory. I argue for a combination of color realism with error: a sophisticated view of color experience allows that colors are real properties, but that naive thought misconceives them. Naive users, though, do count as saying mostly true things about colors. The moral case, I’ll argue, is analogous in some ways and not in others, and the upshot for moral concepts is further from realism than with colors. But as with colors, sophisticates can regard the naive as thinking mostly true things, while perhaps misconceiving what they are doing.

The child who sees a red pillar box as red normally judges it to be red. To accomplish this, note, the child doesn’t need a concept of looking red. It’s we who observe the child who say, rightly, that the pillar box looks red to the child and that this leads the child to think the box red. Soon, though, as the child matures, she can also be convinced that the occasional thing that looks red isn’t really red. A familiar red jacket, for instance, under sodium vapor lights at night may look gray. All this goes for adults too, most of the time in daily life, and perhaps with some adults this is as far as color thinking gets. Young children, and even some adults, may lack an articulate account of what’s going on, beyond that red things may not look red when seen in funny light. But we observers could attribute to them an implicit theory—and it’s not a bad theory. We are detectors of a property that exists independently of us. We detect redness by seeing things as red. The detection isn’t infallible, because sometimes our response isn’t set off by the property, or is set off by things without the property. I call this the naive realist theory of color—not, as I say, because it need be articulated in our naive moments, but because it offers a rationale for our naive judgments and our ways of making those judgments.

In the case of morals, as a parallel might go, what corresponds to seeing as red is an emotional response, calm or more aroused. The naive young child feels outraged at one child’s hitting another, say, and thereby thinks the hitting wrong. Being wrong the child then treats as a property, in that she reasons in terms of wrong in much the way she reasons in terms of a property like being square. At this stage, the child doesn’t need a concept of outrage; it’s we who say that in feeling outraged the child classes the act as wrong. Eventually, though, the child might be convinced that the occasional thing that seems wrong isn’t really wrong, or that the occasional thing that seems all right is really wrong. The same act might seem wrong one hour or one day and not the next. Being wrong is to seeming wrong, then, in this more sophisticated child’s implicit conception, as being red is to seeming red. Whereas seeming red is a matter of color experience.

---

14 Greene et al. (2001).
15 Damasio (1994).
16 For other views, see Boghossian and Velleman (1991); Johnston (1992; 1996).
17 My own view, sketched here, is in my (1996).
seeming wrong is a matter of moral emotions, of feelings of outrage. This we can call naive moral realism. It need not be articulated in our naive moments, but it offers a rationale for our naive judgments. And again, perhaps it’s not a bad theory.

What, then, of guilt? As we have seen, young children lack it, and so, perhaps, do many adults—guilt as a distinct kind of anxiety with its own distinctive properties. For those of us who have it, though, guilt might be another detector of the same property. Writers from guilt cultures indeed often talk of conscience as our prime way of knowing right from wrong, where conscience is the faculty of guilt or self-reproach. Naive moral realism might come, then, in two versions, one narrowly moral and the other quasi-moral. The quasi-moral version has one detector for wrongness, namely outrage, whereas the other has two detectors, outrage on the part of observers and guilt on the part of the agent.

Return, though, to the variant of implicit naive realism that invokes outrage alone. Both with red and with wrong, others too are good but fallible detectors, on this theory. The upshot is disagreement. Moral disagreement is perhaps more frequent and surely more important than disagreement over colors. If Jonathan Haidt is right, it’s primarily interpersonal moral disagreement that prompts moral reasoning. Ordinary reasoning applies to wrongness, since wrongness acts as a property. That’s not to say that reasoning often changes people’s minds, at least when compelling emotions stand behind a judgment. We reason as advocates, not as seekers of the moral truth prone to follow the argument where it leads. Still, we treat reasoning as having the power to persuade—and occasionally it does.18

Mitigated Realism for Color

Should naive realism be our sophisticated view as well? Start again with colors. One naive impression is mistaken: color seems a basic property of things, the kind of salient property one would have to know about to have much comprehension of the real nature of the surface. It’s a surprise, then, that color turns out to be a matter of how uncolored particles interact, and that color is of no explanatory interest apart from the peculiar way that we human beings are constructed. It’s a surprise that a scientist from a planet of Alpha Centauri, if there were one, would rightly take no interest in colors and shades of color, as we delineate them, unless she took an interest in us. It’s a surprise that for subtle questions of color, such as whether two green

ting are the same shade, there may be no fact of the matter; the two may look the same in direct sunlight but different in northern shadow, and if so, there is no intelligible fact of the matter whether they are the same shade or a different shade.

Ought we then to be error theorists about color? Some philosophers think we should be, and if what I have been saying is right, we must indeed attribute implicit mistakes to a naive view of color. But “error theory” as a term of art in philosophy means a theory according to which property ascriptions in a certain realm are systematically false. I point to a British red pillar box and ask the child, “What color is that?” “That’s red,” she answers. An error theory for color would be one that claims that the child is mistaken, that she is saying something false. But didn’t she give the right answer to my question? Didn’t she speak truth? I would say that she did. I would insist, moreover, we rightly take her to be speaking our language, as meaning red by ‘red’.

How should a sophisticated account of color concepts that yields these results go? We know that the straightest of realisms about color won’t work. There’s no fact of the matter precisely what classifications in the world we’re making with our color concepts. There’s no fact of the matter just whose responses in what kind of light are the test of whether two surfaces are the same shade of green. On the other hand, our visual systems do have remarkable properties of color constancy, so that broad aspects of how colors are perceived stay amazingly constant through a wide range of normal conditions. The upshot, I argue, is that color attributions are vague, but admissible resolutions of the vagueness are highly constrained. On any such resolution, colors, color shades, and the like are properties that surfaces really have. These are, as we might put it, objective, physical properties that are of human interest only. Centaurs have no interest in the property of being crimson, and it’s a somewhat vague matter what that property is. But in any resolution of the vagueness, it’s a property of the electromagnetic reflectance of the surface, the degrees to which the surface reflects electromagnetic radiation of various wavelengths, a function from wavelength to proportion of light reflected.19

As for the child who hasn’t thought about vagueness and implicitly treats color as a basic part of the nature of things, we should, as I say, regard her as speaking our language, with certain misconceptions. That’s just a matter of the most charitable interpretation. On any admissible resolution

19 The full story is of course more complex. Reflection from a surface, for instance, must be distinguished from scattering of light in the interior of the material, both of which can give rise to color experience. (I draw this, perhaps in garbled form, from a talk by Mark Johnston some years ago.) I’ll skip over the refinements, however.
of vagueness, the pillar box is red. The child reasons in roughly the ways one should reason if thinking about colors and deploying the most sophisticated conceptions of color. As so interpreted, she gets most matters right and gets some of them wrong—and she might, if she thinks in certain directions, ask certain questions for which there is no clear right or wrong answer. Naive color realism is naive indeed, but it counts as a view about color as color. The right view of color, then, and of the child's color vocabulary, is what we might call a mitigated realism.

Moral Concepts

Can we tell the same mitigated realist story for morals? It might seem so. It has been perennially debated, to be sure, how variable moral judgments are from person to person and from culture to culture, and what the causes of variations are. Moral judgments at their best, it seems, somehow standardize our emotional responses. On the hypothesis that moral sentiments are error-prone detectors of properties—properties of interest to human beings in general or to members of particular cultural groups—we expect some variation in response. We come to views of what the conditions are under which the responses are most nearly constant. We use our responses under these conditions as fallible tests of theories of moral properties.

All this fits in well with a powerful empirical argument mounted by the social psychologist Jonathan Haidt. In his review article "The Emotional Dog and its Rational Tail", Haidt argues that emotional judgment is driven primarily by emotion. Argument matters, he agrees, but except for people who are specially trained as philosophers, argument comes only when the emotional intuitions of different people clash. Typically, each person is convinced that arguments support his own intuitions. "Moral reasoning is rarely the direct cause of moral judgment", he concludes, especially when people have strong feelings on a matter.20 "Reasoned persuasion works not by providing logically compelling arguments but by triggering new affectively valenced intuitions in the listener." Persuasion, when it comes, rarely stems directly from moral reasoning. When people change their minds, morally, on questions they feel strongly about, this is far more likely to be the upshot of new intuitions that arise when the question is "reframed", seen in a new Gestalt.

Haidt's "emotionally valenced intuitions", it seems to me, are just what I call emotions or feelings or sentiments. Feelings are about circumstances and things, over matters, and directed toward people, things, and the like; they aren't just "raw feels" or sensations. As Antonio Damasio describes in Descartes' Error, emotions involve visceral sensations, somehow "bundled" into one's conception of a situation.21 The term "emotion" might suggest high somatic arousal as a part of this syndrome, but I intend it to cover what Hume called "calm passions" as well as "violent passions".

We might, then, try refining naive conceptions of our knowledge of right and wrong into a mitigated realism, a sophisticated view of us as detectors of a vaguely defined property. As with colors, we can try saying, precisely which property is the property of being morally wrong will be a matter of some vagueness. It might be cause for surprise that right and wrong are of interest only to beings like us, and that the gods wouldn't care about these properties unless they were highly anthropomorphic—or even, perhaps, western European in their cultural roots. All this, however, just means that our realism about morals must be mitigated. It doesn't distinguish morals from colors. Is anything wrong with a story like this?

The problem, we expressivists will insist, is that it yields the wrong account of moral disagreement. My wife sees our wood-frame house as a grayish blue, whereas I see it as a bluish gray. Which is it? (Occasionally, in certain lights, I see it her way, and the experience is strikingly different from my usual one. Physical reflectances form a continuum, but color experience doesn't, I note in my own experience.) Well, there's no interesting answer, beyond that, I gather, more people see it her way than mine. We could sharpen our color concepts in various ways, on some of which my house is grayish blue and on some of which it is bluish gray. If she counts as using one such concept and I as another, we're not disagreeing. Do we count as having different concepts? There's no real fact of the matter; the issue of what color our house is lies in the penumbra of vagueness for color concepts. There's no clear matter at issue between my wife and me.

What, then, if we disagree on whether prostitution is always wrong? We might have different feelings about the matter and clashing moral convictions. Is this because there's nothing clear at issue? To say this seems to miss the point. At issue is whether to be against all prostitution, with a special emotional flavor of being against. It's an issue of whether to reprehend all prostitution—or whether to reprehend it if one takes an impartial standpoint of full emotional engagement with the matter.

In saying all this, I do have to allow for an error committed by common sense. We—or most people, in any case—are convinced that our strongest moral views are supported by irresistible arguments, and that if others

21 Damasio (1994). I'm putting this in my own terms, and not making some of the distinctions that Damasio might regard as crucial.
would only listen to reason in good faith, they would be convinced. We also think that if the arguments, to our surprise, went the other way, we ourselves would listen to reason and change our minds. Jonathan Haidt and co-workers have experimented with this latter conviction, describing, for instance, a case of incest where none of the normally expectable harms turn out to ensue. His subjects are disturbed to find all of their arguments convincingly refuted, but they don’t budge on whether the incest was wrong.\textsuperscript{22} Haidt also describes the ordinary conviction that those who aren’t convinced by one’s arguments aren’t in good faith.\textsuperscript{23} If we attend to the evidence, then, we have to be error theorists in one sense: we have to accept that palpable errors infect our ordinary views of morality.

Must we be error theorists, though, in Mackie’s sense? Should we conclude that ascriptions of wrongness, in the ordinary sense, are uniformly false? No more than with color, I’d argue. Just as with a color term like ‘green’, there’s an account of what we could mean by ‘wrong’ that vindicates the core of our ordinary judgments and practices. In the case of color, the vindicating theory is a mitigated color realism that treats red and green as physical properties of surfaces, vaguely determined and of human interest only. It treats our visual systems and color experience as fallible detectors of this property. In the case of morals, I maintain, the story to tell is quite different, but there likewise does exist a vindicating story.

I’ll say a few things to indicate how that story goes, but for any extensive development I must defer to my two books \textit{Wise Choices, Apt Feelings} and \textit{Thinking How to Live}. Questions of moral right and wrong, I propose, are at base questions of how to feel about things people do or might do. Similarly with questions of what one ought to do, in the sense of what one has most reason to do, all things considered: these are questions of what to do. Convictions on these questions consist in something like plans: plans for what to do and plans for how to feel. This view, I argue, though its starting points are like those of Ayer’s emotivism—a paradigm anti-realist view of morality—ends up endorsing many tenets of moral realism. It ends up unclear, then, whether to classify the view as a form of moral realism or a form of moral irrealism. On the one hand, according to the theory, moral concepts are distinct from all naturalistic concepts: wrong can’t be defined analytically in terms suitable for incorporation in psychology or social science. Moral beliefs can be understood as contingency plans of a certain kind, or by the ways they are logically tied to such contingency plans. Naturalistic beliefs can’t be explained in such a way. On the other hand, anyone who plans what to do and how to feel is committed to a

\textsuperscript{22} Björklund \textit{et al.} (2000).

\textsuperscript{23} Haidt (2001).

thesis of property identity: there is a property of being morally wrong. This is an ordinary property, not what Mackie would call a “ queer” property. On any reasonable view, it is, in a broad sense, a natural property—though whether it is a property that should figure in the psycho-social sciences, that cuts our social nature “at its joints” for causal/explanatory purposes, isn’t anything that metatheory can establish. These conclusions would all be accepted by clear moral realists like Sidgwick, Moore, and Ross, and by such current moral realists as Nagel, Dworkin, and Parfit. Whether that makes this metatheory a form of “moral realism” may then be no more than a matter of stipulation, a question of what the term “realism” is to mean.

\textbf{Realism: How Morals and Colors Differ}

Is what I have been saying just an error theory for morals in camouflage? Naively, I have agreed, we implicitly accept a view of ourselves as detectors of an objective property—and that view turns out to be untenable. A sharply different and definable kind of view, I add, matches this naive view well enough that we can, speaking the sophisticated language, charitably interpret ordinary thought. This correct view has many of the earmarks of realism. In particular, moral wrongness, on the sophisticated view, is a property of an ordinary kind. It might, for all the very concept of moral wrongness tells us, be the property identified by a hedonistic rule utilitarianism. Still, the defensible view isn’t the naive view, and so, it seems, the naive view must just be false. It’s false, if what I’ve been saying is right, even if a defensible replacement mimics it pretty well. Does honesty demand a confession like this?

I think not. To see why, turn again to the case of color. The difference between red and wrong isn’t that the naive view of color is mostly right and the naive view of morals is off base. It’s a difference in how the naive view must be refined to be defensible. Both refinements involve surprises. Naively, there seems to be something genuinely at issue in the question of what shade of color a thing genuinely is, apart from how we happen to perceive it. Likewise for the question of who sees things correctly, and under what conditions. All this turns out to be an illusion. With morals, there correspondingly seems to be something genuinely at issue in the question of whether, say, all prostitution is wrong. This time, though, it’s no illusion. The best sophisticated view of morals, if I am right, vindicates this as a genuine issue. The conviction that the issue is real is one we can’t reasonably give up.

What in our conceptions and practices, then, drives the refinements apart? The two naive, implicit property-detector theories, for wrong and for red, seem parallel. Why would their defensible refinements diverge?
The difference stems from contrasts that I’ve so far glided past, contrasts in our naïve conceptions and practices and the points we can find in them. Color responses and emotional responses differ crucially in that emotions are systematically valenced. To reprehend something is to oppose it, to be against it in one’s feelings. Colors too have their valences, from time to time, but not in the same, systematic way. In consequence, the point we can find in thoughts of color is to classify things and to recognize the classifications using our eyes. These classifications are useful for a miscellany of purposes, and nothing much hinges on where exactly we regard the lines between them as lying. What shade of green to see this grass as is not a question of any systematic import. And moreover, how to see the grass isn’t up to us. Feelings of reprehension, in contrast, do have a systematic import. To reprehend, as I say, is to be against. That’s barring refinements, to be sure: in unusual circumstances, I might feel reprehension at something but not position myself against it. Perhaps I thought the matter over yesterday and discussed it with you, and we ended up not opposing it. But we can’t, in any blanket way, divorce feelings of reprehension from being against. Then too, feelings respond to judgments in a way that color experience doesn’t. We can ask ourselves how to feel about something, and the feelings themselves are somewhat responsive to the answer. So when we respond differently, say, to prostitution, there’s a real question to ask: how to respond. There’s something at stake in the answer to this question, namely what to be against.

We start out, then, with similar implicit theories of red and of wrong. At the most naive, the child just regards things as red or wrong, with firm conviction. We observers can note that what leads to these property attributions is, in both cases, the child’s responses: seeing a thing as red in the one case, and feeling reprehension over an action in the other. Indeed at this stage, we’ll see from our eventual, sophisticated standpoint, the child isn’t making any systematic mistake. Most of the things the child thinks red really are red, and most of the things the child thinks wrong really are wrong.

To the theory of what’s red and what’s wrong, the child might then add the beginnings of an epistemology: I know that it’s red because I see that it’s red, and I know that it’s wrong because I apprehend that it’s wrong. (My guess is that children make this move with red but not with wrong: this talk of apprehension sounds not so much child-like as like W. D. Ross.) The property-detector theory this leads to is parallel in the two cases, red and wrong.

Divergence then comes when we ask what could vindicate naïve practice. In neither case does a property-detector epistemology straightforwardly do the job. Color theory needs the realization that the precise boundaries are arbitrary, that there’s no correspondence in how things are colored, apart from us, that isn’t rooted in rough similarities of our color responses. Nothing is systematically at stake in where precisely colors and color shades are to be delineated. With emotional responses, though, something very much is at stake: what to be for and what to be against. Regarding a feeling as warranted or not matters, in that our feelings are responsive to these judgments, and thus what we’re for and what we’re against responds to these judgments. When responses differ, we can’t take the upshot to be mere vagueness in proper classifications, to be resolved arbitrarily if they merit the bother of resolving them at all. We can’t do this and follow our normal bent to use judgments of right and wrong as guides to conduct. Moral classification ties in systematically with what to be for and what against.

In both cases, the best vindication of the main lines of ordinary practice turns out to be different from the straight property-detector view that naïve thought tends toward. In the case of red, it’s a property-detector view mitigated by thoughts of vagueness, arbitrariness, and the like; in the case of wrong, I argue, it’s a view that what’s at stake in morals is how to feel about actions, and thus what to be for and what to be against. Still, though a naïve epistemology is in error in both cases, that doesn’t mean that judgments of red or of wrong are mostly false, or that ordinary modes of investigation are off base. In both cases, a sophisticated vindication of ordinary practice can be found. In both cases, naïve thought fits in closely enough with a defensible explication that charity demands treating naïve thought as mostly getting matters right.

People find the truth about color surprising. If a theory told us there’s nothing to be surprised at, that we get along perfectly well in our ordinary thought without all this high-falutin science of color and philosophical analysis, that would discredit the theory. If I am right about the concept wrong, then like remarks go for morality. The surprising truth, though, isn’t that nothing is truly red or green—and it isn’t that nothing is truly right or wrong.

REFERENCES


Contents

Notes on Contributors vi

Introduction 1

1. Normative Guidance Peter Railton 3

2. Saying what we Mean: An Argument against Expressivism Terence Cuneo 35

3. Expressivism, Yes! Relativism, No! Terry Horgan and Mark Timmons 73

4. Anthropocentric Constraints on Human Value Justin D'Arms and Daniel Jacobson 99

5. The Meaning of ‘Ought’ Ralph Wedgwood 127

6. Knowing Enough to Disagree: A New Response to the Moral Twin Earth Argument Mark van Roojen 161

7. Moral Feelings and Moral Concepts Allan Gibbard 195


9. Direction of Fit and Motivational Cognitivism Sergio Tenenbaum 235

10. Misunderstanding Metaethics: Korsgaard’s Rejection of Realism Nadeem J. Z. Hussain and Nishi Shah 265

11. Resisting the Buck-Passing Account of Value Pekka Väyrynen 295

12. Normativity Derek Parfit 325

Index 381