

## *Mistaken Expressions*<sup>1</sup>

JOSHUA GERT  
Florida State University  
Tallahassee, FL 32306-1500  
USA

It is a suggestive feature of English and other languages that an indicative sentence such as 'Premarital sex is wrong' can be described not only as an expression of the belief that premarital sex is wrong, but also as an expression of disapproval of premarital sex. Disapproval is plausibly regarded as an attitude that is distinct from belief, in that it does not have truth conditions. What sort of attitude, then, should we take 'Premarital sex is wrong' to express: disapproval, belief, or perhaps both? One group of contemporary philosophers advocates the first option. They hold that evaluative claims serve essentially to express positive and negative attitudes that are more like desires than beliefs, and that cannot be said to be true or false — at least in the robust way in which claims about the ages of trees (for example) can be true or false. Call these philosophers 'expressivists.'<sup>2</sup> Seemingly opposed to expressivists are those who hold

---

1 Many thanks to David Copp for some extremely helpful comments on an earlier version of this paper, and apologies for any remaining misrepresentations of his view. Thanks also to an anonymous reviewer for an extremely helpful criticism, and to AlMele and Victoria Costa. Much of this paper was completed with the assistance of the Stephen Risley Family Fellowship, for which I would like to express my gratitude to Stephen and Dottie Risley.

2 See, e.g., Simon Blackburn, *Essays in Quasi-Realism* (New York: Oxford University Press 1993) and *Ruling Passions* (New York: Oxford University Press 1998) and Allan Gibbard, *Wise Choices, Apt Feelings* (Cambridge, MA: Harvard University Press 1990) and *Thinking How to Live* (Cambridge, MA: Harvard University Press 2003). Blackburn has tried repeatedly to argue that there is a sense in which we can properly characterize evaluative claims as true or false. But the story he has to tell is a complicated and controversial one. The present paper offers a simpler way to reconcile the seemingly conflicting intuitions that Blackburn works so hard to render consistent.

that evaluative claims express beliefs, and can be true or false. Call these philosophers 'realists.'<sup>3</sup> A third option has recently been suggested by David Copp, according to which moral claims express both moral beliefs and a certain kind of approval.<sup>4</sup> Copp calls this option 'realist-expressivism,' but for reasons of terminological convenience we can call it 'the hybrid view.'

In arguing against expressivists, the analogy between color and value has seemed useful to many realists. This is true despite debates that go on amongst color-theorists regarding both the ontological status of colors and the objectivity of color claims.<sup>5</sup> For if a realist could manage to show that the claim that capital punishment is wrong is just as reasonably regarded as objectively true as the claim that snow is white, he might well regard his work as done. He need pay little regard to the philosopher who subsequently asks how reasonable it is to regard it as objectively true that snow is white. Against this sort of defense of realism, some philosophers have explicitly attacked the analogy between color and value.<sup>6</sup> The purpose of this paper is to suggest that the analogy can actually serve to reconcile expressivists and realists, at least for some very central normative terms, yielding a view that shares some features with Copp's hybrid view.<sup>7</sup> However, unlike Copp's view, the view developed here makes the relevant evaluative discourse much more like normal descriptive discourse. Moreover, while it is essential to expressivism that a sincere and competent speaker who makes an evaluative claim has a certain pro- or con-attitude, and while Copp's view involves the idea that something needs explaining whenever such a speaker does

---

3 See David Brink, 'Externalist Moral Realism,' *The Southern Journal of Philosophy* 24 (1986) 23-42 and Peter Railton, 'Facts and Values,' *Philosophical Topics* 14 (1986) 5-31.

4 David Copp, 'Realist-Expressivism: A Neglected Option for Moral Realism,' *Social Philosophy and Policy* 18 (2001) 1-43. See also Oswald Hanfling, 'Learning about Right and Wrong: Ethics and Language,' *Philosophy* 78 (2003) 25-41.

5 For a useful collection of relevant papers, see Alex Byrne and David Hilbert, eds., *Readings on Color, Vol. 1* (Cambridge MA: MIT Press 1997).

6 See Blackburn, 'Errors and the Phenomenology of Value,' reprinted in *Essays in Quasi-Realism* and Crispin Wright, 'Moral Values, Projection and Secondary Qualities,' *Aristotelian Society Supplement* 62 (1988) 1-26. Wright, however, does not attack the analogy in order to defend an expressivist view.

7 The availability of the present argument for this reconciliation therefore depends on the falsity of extreme subjectivist views of color such as C.L. Hardin's. See, e.g., C.L. Hardin, 'A Spectral Reflectance Doth Not a Color Make,' *Journal of Philosophy* 100 (2003) 191-202. This should not worry readers who are willing to grant, for example, that there are such things as red roses and blue violets.

not have the relevant attitude, the proposal advocated here does not have this strong implication.<sup>8</sup> Rather, many evaluative claims made in the absence of the relevant pro or con attitude need be no more exceptional than color claims based on testimony or memory.

## I Agreement, Ostension and Reference

We human beings are so constituted that that we have more or less the same phenomenological responses to variously colored objects seen under the same circumstances. Even those who worry about inverted spectra have to concede that human beings make virtually the same similarity judgments regarding the colors of different objects.<sup>9</sup> And this agreement in similarity judgments is all that is needed for a community of language speakers to make use of ostensive methods to teach language learners the meanings of common color words such as 'green' and 'orange.' For it is this agreement that allows for the following two things. First, we can be pretty sure that the overwhelming majority of language learners will go on in the same way as we do after the requisite number of ostensive uses have been made. Second, those few humans who are not set up to go on in the same way (color-blind people, for example) will be corrected when they go on in the wrong way. Perhaps they will never, therefore, be able to apply color words correctly simply by looking at objects. But they will learn that there are acceptable and unacceptable way of using those words, and that they can generally trust other people to tell them the colors of objects. Thus the overwhelming agreement in phenomenal response is sufficient to explain why the meanings of color words are such that it is typically either true or false that a given object has the color that such a word picks out.<sup>10</sup> Of course, the brute fact of the matter is merely that ninety-seven percent of the population agrees in

---

8 See Copp, *Realist-Expressivism*, 17-18.

9 Interestingly, because color experience results from a dimension-lowering process that takes (roughly put) three input channels to two output channels, and because the three input channels can be calibrated differently in different normal humans, many pairs of colored patches that are indistinguishable to me may be very easily distinguished by you, and vice versa, without any deficiency or abnormality in either of us. Thus 'similarity in similarity judgments' is the best we can do in this claim about shared human responses. But it is sufficient.

10 'Typically' on account of the vagueness inherent in color words.

ninety-seven percent of the cases.<sup>11</sup> Someone who focuses too narrowly on this sort of brute fact might therefore advocate a kind of ‘visual liberation,’ according to which the minority has as much a right to their views as the majority. But this would be a mistake. Given the overwhelming agreement in the community, even the minority is constrained to learn the meanings of color words in line with the way those words are taught. And those words are taught by ostension.<sup>12</sup> If those who teach the language to the next generation of language speakers would unhesitatingly use a certain object to teach the word ‘red,’ then that object is an appropriate one for the teaching of the word ‘red.’<sup>13</sup> In this way ostensive teaching parlays ‘mere’ overwhelming agreement in phenomenological response into a more complete agreement in meaning — even among people who do not share the relevant phenomenological response. For even those with visual problems are constrained to use the word ‘red’ with the univocal meaning it has in the language.

It is possible to regard the preceding remarks concerning the meaning of color terms as advocating a so-called ‘response-dependent’ account of color: an account according to which the meanings of color-words are semantically linked to the responses of human beings. This may be true, but if it is, it should be noted that it is a relatively uncontroversial response-dependence. For example, it is not the (false) view that ‘red’ means ‘such as to produce the same sort of phenomenological experience that you are having now [as I hold up a red object] in the overwhelming majority of people.’<sup>14</sup> For what is important to the meaning of a word is whatever people learn when they learn the meaning of that word, and the fact of overwhelming agreement is simply not part of what people learn when they learn color words. This is easy to see when one realizes that even if there were not overwhelming agreement in the visual responses of human beings, a parent who shared these phenomenological responses with his child could teach his child a word with the same

---

11 This agreement should be understood as agreement about whether or not two colored objects are quite similar, quite different, and so on.

12 By ‘ostension’ I mean to indicate more than the paradigmatic instance of ostensive teaching, which is rather rare. I mean to include also the sort of language learning to which the following is essential: the object, stuff, or property for which a word is being taught is present to the learner in a way that would also allow paradigmatic ostensive teaching to happen.

13 This is true even if the object is not actually red.

14 For a view of value that does take this implausible form, see David Lewis, ‘Dispositional Theories of Value,’ *Proceedings of the Aristotelian Society*, Suppl. Vol. 63 (1989) 113-37.

meaning as 'red.' The role for overwhelming agreement is the following: it allows for the teaching of color words to language learners at large, by language teachers at large, so that the word 'red' in English has the objective character it does: the same objective character that a similar word might have had in the tiny idiolect of the father and child mentioned above.

Nor does the view of the meanings of color-words suggested by this paper involve the controversial sort of response-dependence that claims that the concept of redness is dependent on the concept of our responses.<sup>15</sup> For the account put forward here is simply an account of how human beings learn to use color words in the right way — through ostension and correction. It makes no claims about how or whether this ostension and correction involves prior concepts. All that is being claimed is that the overwhelming agreement in phenomenological response is part of an explanation of the fact that 'red' is a referring word. If this is right, then the word 'red' refers to the class of things that can be described (very roughly) in the following way: things to which the overwhelming majority of people have the same phenomenological response as they have to this [ostending a red object]. But the contingent fact that this description happens to capture the referent of the word 'red' says very little about the relation between this description and the meaning of 'red.' For the referent might also be described in the following way: things that have a reflectance-pattern in the set  $\pi$ .<sup>16</sup> And one can understand the meaning of 'red' while knowing nothing of the wave-theory of light.

## II Expressive Meaning and Reference

Because the word 'red' has an objective referent, many ascriptions of redness to objects have truth-values. This explains why it is that when a competent English-speaker calls an object 'red,' it is possible to understand this as the expression of a belief. But it is also possible — in many cases — to understand color claims as expressions of (not reports of) a

---

15 Mark Johnston, 'Dispositional Theories of Value,' *Proceedings of the Aristotelian Society*, Suppl. Vol. 63 (1989) 139-74.

16 In fact, this latter claim is false, but is meant to do duty for a more accurate but prohibitively complex physical specification of the referent of 'red.' See C. L. Hardin, *Color for Philosophers: Unweaving the Rainbow*, expanded edition, Indianapolis: Hackett Publishing Co. 1993: 7, 62, 64.

relevant phenomenological experience.<sup>17</sup> After all, it is the fact that the speaker is having that experience that often explains why the color ascription gets uttered. This latter understanding of color-ascribing assertions is not typical, and may not be useful for many purposes. But it seems at home when we imagine very young children who are just learning to make color distinctions, and who are delighted to express what they are experiencing. And it also seems at home in the context of experiments with adults who are told to make quick color judgments when a patch of color is flashed in their visual field.<sup>18</sup> But in any case, there is no need to separate instances in which a color-ascribing assertion is 'really' the expression of a belief, and instances in which such an assertion is 'really' an expression of a relevant phenomenological experience. For, as will also be claimed for evaluative assertions, the very same utterance can often be usefully understood as the expression of more than one attitude.

It is no accident that the phenomenological experiences that color ascriptions can be understood as expressing are the very same experiences that are instrumental in explaining how color words get their objective referents. For language-learners are taught to make the ascriptions precisely when they are having the relevant phenomenological experiences: that is, they are taught how to identify the colors of objects simply by looking at them. And yet, even when someone who has learnt the language is actually having the right sort of phenomenological experience, and expresses this by some color ascription, the possibility remains that the ascription may be mistaken. For example, the person may be unaware that a strong after-image has distorted his color-perception, or that the object is bathed in a colored light. This possibility of mistake leads to a problem in deciding how to describe the meanings of color words. The fact that the person is sincere and competent in his ascription suggests that the meaning of a color word is given by its role in expressing a certain sort of phenomenological experience. For in such a case the person understands the meaning, and his use manifests that understanding perfectly: the speaker has not misused the word, simply

---

17 Compare Frank Jackson and Philip Pettit, 'A Problem for Expressivism,' *Analysis* 58 (1998) 239-51. Their point tends in the opposite, but perfectly compatible direction: that expressions of attitude can be understood as reports that one has an attitude.

18 Compare Ludwig Wittgenstein, *Philosophical Investigations*, trans. G.E.M. Anscombe (Upper Saddle River, NJ: Prentice Hall 1973): 'Both things, both the report and the exclamation, are expressions of perception and of visual experience. But the exclamation is so in a different sense from the report: it is forced from us. It is related to the experience as a cry is to pain' (197).

because he has made an error regarding the color of the object. Indeed, it is the assumption that the speaker has not misused the word that allows us to infer that there is something distorting the appearance of the object that she is describing. Nevertheless, the fact that the person is mistaken suggests that the meaning of the color word is given by its objective referent. For it is the presence or absence of the referent that determines the truth-value of the sentence the speaker utters. Which then should we choose, the 'expressive meaning' or the 'referring meaning'?

In fact, we need not choose only one. It is a philosophical commonplace that the referent of a linguistic expression may be different from the speaker's referent in using that expression. Illustrations of Gettier's problem typically involve such differences.<sup>19</sup> For with the phrase 'the person in the corner with the martini' the speaker may be referring to Smith, who is in the corner drinking water in a martini glass, while the phrase itself refers to Jones, who is drinking a martini in a coffee cup. Such examples show that this difference in reference does not depend upon the speaker's lack of competence with the relevant concepts, or his lack of sincerity. Gettier offered his examples in the context of a discussion of the definition of knowledge, and Donnellan offered similar examples in the context of a discussion of reference, so it no surprise that such examples always involve distinctions between speaker's reference and word or phrase reference. But there is nothing in principle that requires us to hold that the only difference that appears when we shift our attention from the speaker to the phrase is a difference in reference. If the function of a word in the mouth of a particular speaker is to express a phenomenological experience, then that word can be taken to have an expressive speaker's meaning on that occasion. But if the same word in the language at large serves to pick out a determinate referent, then it may not have an expressive meaning in the language, but a referring one.

Now it is also a philosophical commonplace that much of the information communicated by an utterance need not be regarded as any part of the meaning of the utterance: not even a part of the speaker's meaning. If someone is asked her opinion regarding the philosophical talent of a colleague, the sentence 'He's a great person, really sweet,' given in answer, need not be taken to mean, even by the speaker, that her colleague lacks talent. Rather, a speaker offers this answer knowing that the person with whom she is speaking will make a certain inference, based on a number of conventions that govern conversational ex-

---

19 See Edmund Gettier, 'Is Justified True Belief Knowledge?', *Analysis* 23 (1963) 121-3; see also Keith Donnellan, 'Reference and Definite Descriptions,' *Philosophical Review* 75 (1966) 281-304.

changes. Indeed, it is because the speaker knows what her answer means, and that it is not directly relevant to the question of philosophical talent, that she can use this answer to communicate a further attitude. This sort of conversational implication is very common.<sup>20</sup> It would not be very interesting if the expression of a pro or con attitude that one made in uttering an evaluative claim were accomplished merely through this sort of implication. For virtually every realist would be willing to admit that, typically, when one utters an evaluative claim, one can reasonably expect that one's audience will assume that one has the relevant sort of attitude. If this admission were sufficient to turn one from a pure realist into an advocate of a hybrid view, then there would be very few pure realists. The view defended in this paper is not this relatively uninteresting view — a view that even pure realists would endorse. Rather, the connection between expressions of attitudes and evaluative utterances is in some cases tighter than mere conversational implication: it is the same connection that exists between color-ascriptions and experiences of color. This latter connection has, in certain circumstances, a species of necessity. In particular, at least in certain circumstances, someone who makes a color-ascription cannot cancel the implication — that is, cannot sensibly deny — that she is having a certain phenomenological experience. For in certain circumstances we must assume that the color ascription *is to be explained* in part by reference to the fact that the speaker is having the right sort of phenomenal experience. Now, it will turn out that the same sort of claim cannot be made in the case of evaluative utterances. Nevertheless, this fact can be explained without denying that the relation between utterances and attitudes is the same, in relevant respects, as in the case of color.

### III A Common Structure in Color Ascriptions and Evaluative Claims

Consider again the story that was told about color words. The idea was that an overwhelming agreement in the phenomenological responses of human beings is what allows for the ostensive teaching of these words. This agreement allows even color-blind people to learn that there are facts of the matter with regard to the colors of objects. Such people cannot learn the colors of objects simply by looking at them, but they can

---

20 See Paul Grice, *Studies in the Ways of Words* (Cambridge, MA: Harvard University Press 1989) 22-57.

nevertheless, via other means, learn that certain objects have certain colors, and they can use this knowledge to describe things, ask for them, and so on. However, a competent language speaker without any visual problems can often tell the color of an object simply by looking at the object, and given no other clue as to its color. In a case of this sort, the inference one draws from a color-ascription — the inference that the speaker is having a certain phenomenological experience — is not the result of mere conversational implication. It is not cancelable, for example, simply by denying the claim that otherwise would be implied. What would we make of a person, identifying a series colored patches, whose identifications took the following form: 'Blue, though it doesn't look blue to me'; or 'Red, though it doesn't look red to me'? If the person thought that she was suffering some systematic distortion in color perception, or that some colored light was obscuring the true color of the object, we could perhaps understand this. But let us stipulate that the speaker does not take herself to have reason to suspect any such distortion. In such a case, sincere color claims cannot be accompanied by associated denials of phenomenology without abuse of language. So the implication, whatever it is, is not conversational, and in some cases is a strict implication.

It may seem that I have filled out the description of the situation in such a way that we could draw the 'inference' from the speaker's claim without actually making any reference to anything that the speaker says. That is, if I have stipulated that the speaker has no problems with her vision and is looking at a blue object under circumstances that allow an accurate perception of its color, it seems that I can infer that she is having a certain sort of phenomenal experience, regardless of what she says. Therefore, one might suggest, this inference reveals nothing about the expressive nature of color claims: the fact that the person makes an utterance is an idle part of the scenario. But the stipulation was not that the speaker had no visual problems and that no distorting light was affecting the appearance of the object. Rather, it was that the speaker did not take herself to have any reason to suspect any such things. That is, the speaker is in a position to say 'My visual system is functioning well, and my circumstances for viewing this object do not include any distorting factors, and the object is blue.' In such a circumstance, we would not know what to make of the further claim 'but nevertheless, I am not having a experience as of a blue object when I look at the object.' This continues to be true even if we happen to know that the object is green, and even if we have no idea why the person is misperceiving it.

If one has a wrong view of the way in which we manage to make color ascriptions, it may seem more plausible that the implication from a sincere and correct color ascription to a certain phenomenological experience is merely conversational. The wrong view is that when one makes claims about the colors of objects, one does so by inspecting the color of

the sense datum that the object produces. This view suggests that what one 'really' sees is something in one's internal visual field, and that one has learned to infer from the color of this internal object to the color of the external object. But this view, though it has historically tempted many philosophers, is widely, and correctly, regarded as mistaken. It is the picture attacked in Wittgenstein's private-language argument, and it depends, for its coherence, on the mistaken view that one can meaningfully identify and label one's own internal private experiences and then go on to discover that these experiences are correlated with what other people identify by, say, color words.<sup>21</sup> In contrast to this picture, the view of color ascriptions advocated in this paper is that people learn to identify the colors of objects in practice, and that part of this learning might be characterized very crudely in the following way: one learns to utter 'red!' when presented with something red. This is why color claims can be regarded as expressions of one's experiences, as well as reports about the properties of external objects. One's phenomenological experience plays a mediating role in the training required to learn color language, and in subsequent color ascriptions, but not because one first learns to identify 'the colors of one's experience,' and then learns to ascribe colors based on the fact that there is a correlation between the colors of one's experiences, and the colors of external objects. The correct picture is one on which it makes more sense to describe a color ascription as a verbal expression of a phenomenological experience, and neither a reporting of it, nor a claim one makes based on a prior, private, color ascription. In fact, of course, many of our color ascriptions are not like this, since we base those ascriptions on, for example, testimony or memory. But it remains true that some color ascriptions can be viewed as the expression of a certain experience. In those cases, the connection between the ascription and fact that the speaker is having the experience is much tighter than a mere conversational implication.

The preceding remarks were intended to defuse a worry about the nature of the implication from certain color ascriptions to the fact that the speaker is having a certain experience. The worry was that this implication was mere conversational implication, and had little direct relevance to the semantics of color claims. But the very remarks offered in order to defuse this worry may give rise to another worry. For it may seem that if our color claims cannot be regarded as reports that we are having certain phenomenological experiences, or as claims based on the prior ascription of a phenomenal property to a private experience, then

---

21 See Wittgenstein, *Philosophical Investigations*, §§243-90.

the relevant experiences must be regarded instead merely as the causes of such claims. Indeed, my brief sketch of the learning of color terms supports this idea, for the picture there was that in learning color words we are first of all trained to utter the appropriate words when having the corresponding experiences. But if the relation is merely causal then the claim that there is any implication seems endangered. After all, the claim that there is some sort of implication is that there is something semantic going on that depends on linguistic conventions. The answer to this objection is that a sharp distinction between causal and semantic relations cannot be drawn, and is especially problematic in cases of first-person ascriptions of psychological states. Consider pain. It naturally causes wincing, crying, and other pain-behaviors. We can therefore often reasonably infer from wincing and crying to the fact that the person who is wincing and crying is in pain. This doesn't mean that there is any implication — conversational, conventional, strict, or loose — from the wincing and crying to the fact that the person is in pain. Part of the reason there is no such implication is that wincing and crying do not have any content: they are not pieces of linguistic behavior. But on one very plausible view of first-person pain ascriptions, claims such as 'That hurts' or 'I'm in pain' are usefully understood as modifications or extensions of natural pain behavior. They are learned behaviors that replace wincing and crying.<sup>22</sup> And yet, there is an implication from such first-person claims to the claim that the speaker is in pain. How can this be, if the only difference is that wincing and crying are natural expressions of pain, while claims such as 'That really hurts' are learned expressions of pain? The answer is that this difference is precisely what matters. In learning the verbal expressions of pain, one learns how to express pain in a way that meshes with our other linguistic abilities, so that one can respond, verbally, to queries such as 'Where does it hurt?' and 'How much does it hurt?' At this point, and in this linguistic form, pain behavior begins to have content. And it is this that allows us sensibly to talk about the implications that such claims have. Now, I have claimed that 'That object is blue' is sometimes the expression of a phenomenological experience in exactly the same way in which 'That hurts' is an expression of pain.<sup>23</sup> The link may be causal, but it is also semantic.

---

22 Compare Wittgenstein, *Philosophical Investigations*, §§244-5.

23 Is 'That hurts' also an expression of the belief that one is in pain? Wittgenstein himself might be inclined to deny this, since he sees little sense in a sentence such as 'I believe I am in pain,' and also clearly distinguishes the criteria for third-person ascriptions of pain from first-person ascriptions. See Wittgenstein, *Philosophical Investigations*, §§251 and 288. See also Ludwig Wittgenstein, *Remarks on the Philosophy of Psychology*,

This does not mean that every cause of a linguistic utterance has a semantic link to the content of that utterance. But if the particular cause will figure in a plausible description of what counts as an acceptable use of the relevant term or phrase, then the link is just as plausibly a semantic one.

With these worries about the expressive nature of color claims behind us, let us now consider a normative word such as 'harm.' It is plausible that this word developed in the language because there was (and continues to be) a more or less fixed class of events — suffering pain, dying prematurely, losing physical abilities or freedoms — to which the overwhelming majority of human beings share a common phenomenological response: some type of aversion. This overwhelming agreement in response may be what allows the word 'harm,' and similar words, to be taught in an ostensive way, as color words are. This allows for sincere and competent, but mistaken applications of the word 'harm.' For example, it explains how it might be true that although someone sincerely regards the loss of his inherited title as a harm, and expresses this by claiming that it is a harm, in reality it is not. Now, there may be people who are not averse to certain harms, just as there are people with defective responses to the colors of objects. We may wish to call these people 'irrational' in some respect, just as we call those with defective responses to color 'color-blind.'<sup>24</sup> These people, whatever we call them, can nevertheless form justified beliefs about the harmful nature of certain things. However, if there is no reason to suspect that a speaker's aversive responses are distorted, then, just as in the case of color, the speaker can express justified beliefs about the harmful nature of things simply by being presented with them — not based on any inference. This explains how certain evaluative claims can both have truth conditions of the same sort that color claims have, while also serving to express attitudes such as aversion.

---

Vol. 1, G.E.M. Anscombe and G.H. von Wright, eds.; G.E.M. Anscombe, trans. (Oxford: Basil Blackwell 1980) §§502-4.

24 'Color-blind' and 'irrational' are here functioning as what might be called 'remainder terms.' In the case of the notion of harm, this use of 'irrational' seems plausible. But this need not be true for all normative notions. For example, it is a controversial — and, in my view, false — claim that those who fail to respond to *moral* demands are irrational. Rather, they are immoral.

#### IV A Relevant Difference Between Color Ascriptions and Evaluative Claims

The preceding argument placed a good deal of weight on an analogy between color concepts and evaluative ones, and on a story about how these concepts are linked with words in public languages. In the case of color, one claim was that when a competent language speaker who takes herself to be without any visual problems says what color an object is simply by looking at it (and therefore without any other clue as to its color) we can infer that the speaker is having a relevant phenomenological experience. This is because we know that in such circumstances the speaker's utterance is in part caused by the experience, and can in fact be understood as an expression of that experience. It might seem that these claims, taken together with the analogy between color and value, entail that similar claims are true in the case of evaluative claims. That is, it might seem that I am committed to the idea that when a competent language speaker who takes herself to be without any motivational problems says that a certain thing counts as harmful, simply by considering it, we can infer that the speaker is averse to the thing because we must understand the person to be expressing her aversion by means of her utterance. Call this the 'expressive inference.' While some philosophers would find such a claim unproblematic, part of the reason I reject both expressivism and Copp's proposal is that they imply that in a context in which an amoralist — someone who takes herself to have moral beliefs but who is unmoved by those beliefs — is simply asked his moral opinion (perhaps on a questionnaire) she would be misusing moral language in giving a simple answer.<sup>25</sup> This is not the place to argue whether or not this really is a problem for expressivists or for Copp. Rather, the point of the current paper is only to present a realistic view of at least some evaluative concepts that is consistent with a central intuition that may lead some philosophers to adopt expressivist views:

---

25 It should be clear why expressivism has this implication. Copp's proposal has it for the following reason. Copp holds that moral terms have 'coloring.' One of the conditions for a term's having coloring is that if one uses the term in a sentence such as 'Premarital sex is wrong,' and one knows that one does not have the appropriate attitude, then either there is some pretense involved, or one is misusing the language. Copp calls this the 'misuse' test for coloring. See Copp, *Realist-Expressivism*, 17-20. Copp's view is subtle, however, and it should be noted that the misuse involved need not prevent an utterance from being literally true. Indeed, 'misuse' is probably too strong a term; Copp suggests 'linguistic gaffe,' 'semantic offense,' and 'violation of semantic proprieties' — each of a variety that does not touch truth value.

the intuition that, in some sense, voicing an evaluative belief can be, in itself, the expression of a pro- or con-attitude. Once one has adopted an expressivist view, one may be led to further claims, such as the claim that the amoralist above is abusing the language. But it is no part of my project to offer a view consistent with every aspect of expressivism, or even with every positive claim made by expressivists.

So far I have argued that the connection between certain evaluative claims and certain pro- and con-attitudes is of the same sort as the connection that holds between color ascriptions and certain phenomenological responses. In both of these cases there are circumstances in which the expression of a belief is also the expression of an attitude (or, in any case, a mental state) distinct from belief. What I wish to do, in this section, is to explain how the analogy between color and value need not make the expressive inference legitimate in the case of certain values, even though the same sort of inference may often be legitimate in the case of color. The illegitimacy of the expressive inference in the case of these evaluative claims of course does not mean that such evaluative claims do not, often, express pro- and con-attitudes: surely they often do. It only means that we cannot infer, in the same way, from certain expressions of a speaker's evaluative beliefs, to the fact that that speaker has any particular attitude other than the relevant belief. But of course we might be able to infer in some other way.<sup>26</sup>

Despite the fact that it is the overwhelming agreement in phenomenological response that allows for the development and teaching of objective color-language, I have noted that not everyone who learns color words need participate in this agreement. People who are color-blind can learn color words, and can learn to apply them correctly. What they cannot do is learn to apply them correctly based simply on their own visual experiences. Some have denied that such people 'really' have color concepts.<sup>27</sup> The problem with such a view is that it seems to pick out, without any defense, one privileged mode for identifying color, and to elevate that mode to the status of an essential ability — an ability that anyone who possesses color concepts must have.<sup>28</sup> And the problem with

---

26 The next section explores this possibility for a certain class of evaluative terms: ones that have a conceptual tie to rationality.

27 See Michael Smith, *The Moral Problem* (Cambridge, UK: Blackwell 1994), 51, 69-70. See also Christopher Peacocke, 'Colour Concepts and Colour Experience,' in Byrne and Hilbert *Readings on Color*, 54.

28 See D.M. Armstrong, 'Smart and the Secondary Qualities' in Byrne and Hilbert, *Readings on Color*, 43.

such a move is that it is based on a picture of how we identify substances or properties that fails to apply generally. Generally we say that someone knows what a certain substance or property is when she has the ability to identify that substance or property.<sup>29</sup> And it doesn't matter how she manages to make that identification. A blind person can identify John Smith by voice, by knowing his name, by the feel of his face. Another friend of John Smith can identify him by sight, by handwriting, and so on. The ability to identify John Smith comes in different forms, and in different degrees. There is perhaps a degree below which we would not want to say that someone knew who John Smith was. But in the case of color, it seems fairly clear that a color-blind person has enough ways to identify the colors of objects, and can use that information in enough of the ways that normally-sighted people do, that we should say that such a person knows what colors are, and knows what she is saying when she says that tomatoes are better when they are red.

In the case of evaluative properties such as being harmful, there are also people who fail to participate in the overwhelming agreement that is responsible for the development and teaching of objective evaluative concepts. But such people can, nevertheless, form and express justified beliefs about the harmful nature of various things. Indeed, it is rather easier for such people to form these beliefs than it is for people who are color-blind to form such beliefs about the colors of objects. This is because the most basic access humans have to the colors of objects is via sight. The submicroscopic features of objects, which might at least conceivably be used to determine their colors independently, are simply not available for direct observation by any other faculties. Another way of putting this is that the naturalistic features on which colors supervene are not directly observable. But the naturalistic features on which harm supervenes are directly observable. That is, it is not some indiscernible feature of being jailed — some invisible radiation inside the cell — that makes it harmful, and that we detect via a faculty that produces aversion. Rather, we are simply averse to such limits being placed on our physical freedom. So a person who does not have the standard motivational responses to harms might nevertheless be able to gather, through the processes of language learning, that restrictions on physical freedom count as harms, even though he himself is not averse to such restrictions. On the other hand, a blind person could never learn, in a way that was useful for everyday tasks of color-identification, that spectral reflectance  $\pi$  counted as orange. Because the features on which harmfulness super-

---

29 This view of substance concepts is persuasively defended by Ruth Garrett Millikan in *On Clear and Confused Ideas* (New York: Cambridge University Press 2000).

venes are easily and independently perceptible, there is an important difference, relevant to the legitimacy of the expressive inference, between color ascriptions and evaluative claims. In the case of color ascriptions, the expressive inference is legitimate in cases in which a competent language speaker who does not take herself to have any visual problems, and does not take there to be anything in the environment that is distorting the color of the object, says what color the object is simply by looking at it. In such cases the phrase ‘simply by looking at it’ entails that there are no other clues as to its color available to the speaker. So we can infer that it is the phenomenological experience that is gaining expression in the color ascription: the expressive inference is legitimate. But in the case of evaluative claims, the phrase ‘simply by considering it’ cannot possibly entail that one’s aversive response is the only clue as to — for example — the harmful nature of the thing considered. This is because the subvening properties (that the thing causes pain, is life-threatening, etc.) are also available. So the inference from a speaker’s claim that the thing is harmful to the fact that the speaker is averse, mediated by the idea that the claim expresses this aversion, is never legitimate. It is always possible that the speaker took his clue from the subvening properties, and did not simply express his aversion with an evaluative claim.

## V Expressivism and Internalism

It has just been argued that we can never make the expressive inference in the case of the normative term ‘harm,’ although we can sometimes make an analogous inference in the case of color terms. One might worry that this is plausible only because ‘harm’ is a so-called ‘thick’ evaluative term: one that has both a descriptive and evaluative element. But one might suspect that the same sort of argument will not be plausible for ‘thin’ terms, which are purely evaluative. In the case of thin terms, one might hold, only an expressive analysis will suffice. Of course it is an open question whether there are any genuinely thin normative terms. But one might try to argue for their existence precisely by offering examples in which the expressivist inference seems as legitimate for a certain normative term as the analogous inference does for color terms. This would undermine a response-dependent realistic account of the candidate ‘thin’ term, since, on such an account, the expressivist inference should not be legitimate. Suppose, then, that my friend Michael makes the following claim: ‘I ought to X, and I have not the slightest motivation to X, but also I do not suffer from the slightest degree of irrationality.’ Doesn’t the argument of this paper suggest that this claim ought to be perfectly intelligible, since Michael could be basing his claim

that he ought to X on the relevant subvening naturalistic properties, rather than expressing any particular motivational attitude towards X-ing by means of his utterance? But, contrary to this suggestion, isn't Michael's claim in fact incoherent?

There are two possible responses to this worry. The first is simply to concede the point in the case of 'ought,' but to deny that this concession has any great importance. Nothing in this paper is incompatible with the truth of expressivism for some normative terms. In fact it seems to me rather plausible that 'ought,' both in its rational and moral incarnations, is best given an expressivist semantics. There is no inconsistency in arguing that certain normative notions, such as 'harm' and 'rational' should receive a realistic response-dependent analysis, while others, such as 'ought' or 'beautiful,' should receive an expressivist one. I have argued in a number of other places that it is a mistake to think that all normative terms will yield to the same sort of analysis.<sup>30</sup> The point of this paper has not therefore been to argue that expressivism is universally false. Rather, it has been to explain why and how a realistic semantics for *some* central normative terms is plausible, and why, given such an analysis, we can still see some truth in the expressivist's mode of conceiving the relation between the use of those terms and certain non-cognitive attitudes.

A second response to the worry also concedes that Michael's claim is incoherent, but argues that even if it is, this fact does not — even for the term 'ought' — undermine the account offered in this paper. For the incoherence of Michael's claim might well depend, not on the truth of expressivism for the term 'ought,' but on the truth of a much weaker view, for which we can appropriate the familiar label 'internalism.' Our version of internalism should be understood as a thesis that can be true of some normative terms, and false of others. It holds only that if a sincere, competent and rational person uses the relevant term to make an evaluative claim, then that person must have a relevant non-cognitive attitude. Expressivism, on the other hand, as a general thesis about the semantics of normative language, makes the much stronger claim that

---

30 What may make them all normative terms might be the link to affective attitudes, and what may make some realistic, and others not, might have to do with the degree of agreement in response among human beings. In 'Cognitivism, Expressivism, and Agreement in Response,' which should be appearing in the second volume of *Oxford Studies in Metaethics: Volume 2*, Russ Shafer-Landau, ed. (New York: Oxford University Press, forthcoming), I argue that as agreement in response becomes less universal it becomes increasingly important to know what response the speaker has. This explains why the expressive function of the related word becomes more important, and may eventually result in its becoming essential to correct use.

sincere and competent evaluative utterances serve essentially to express non-cognitive attitudes. As I will now argue, the realistic response-dependent account of normative terms sketched in this paper is perfectly compatible with internalism for certain normative terms, including, plausibly, 'ought.'

It is easy to see the difference between the expressivist and internalist theses by considering the evaluative notion of rationality itself. Let us therefore begin by stipulating, just for illustrative purposes, that 'rationally required' has a realistic, and not an expressivist, semantics. Even with this stipulation, it will turn out that no rational person will be able to sincerely and competently claim that a course of action open to her is rationally required and yet fail to be motivated to perform it. That is, the internalist thesis will be true of the notion of being rationally required. Why? Simply because people who are rational are people who (at the least) perform rationally required actions, and in order to perform them, they must be motivated to do so. This can be true even if the relevant motivation does not find *expression* in the agent's claim that the action is rationally required. For the claim to be an expression of such a motivation, the motivation must bear the right kind of causal relation to the utterance. And the mere truth of internalism in the case of claims about rational requirements provides no reason to think that the relevant motivation does bear such a causal relation. This example therefore shows how the internalist thesis might well be true for an evaluative term, even if the expressivist thesis is false with regard to that term.

The argument just offered for distinguishing expressivism from internalism made use of the notion of 'rationally required.' But it might just as effectively have made use of any other evaluative notion, as long as that notion had the right sort of tie to rationality, and as long as we continued to give rationality itself a realistic analysis.<sup>31</sup> For example, it is very plausible that it is irrational to fail to be averse to harms.<sup>32</sup> If this is right, then no rational person could sincerely and competently claim that a certain consequence of his action counted as a harm, and yet fail to be averse to that consequence. Thus internalism will be true of claims about harms. But again, the truth of internalism in this example does not imply the corresponding truth of expressivism, and for the same reason as in the original argument. For even though the speaker, if rational, will

---

31 I give rationality such an analysis in *Brute Rationality* (Cambridge, UK: Cambridge University Press 2004).

32 Aversion to a particular harm is compatible with a rational intention to act in a way that will bring one that harm. For it may be necessary to suffer the harm in order to avoid a greater harm, or to gain some great good.

be averse to the consequence he regards as a harm, this by no means entails that his utterance *expresses* this aversion.

The general conclusion here is that if a realistically interpreted normative property is one towards which a certain attitude is rationally mandated, the internalist thesis will be true with regard to ascriptions that of that property. Such terms very plausibly include 'rationally required' and 'harmful.' What about 'ought'? The important question here is whether we are rationally required to behave as we ought. The answer to this question is that there are many senses of 'ought' (or, perhaps, many different domains in which the word applies). If we understand 'ought' as 'rationally ought,' then it is exceedingly plausible that we are rationally required to behave as we ought. Because of this, the internalist thesis is plausibly true of this rational sense of 'ought.' But there is also the moral ought. Understanding 'ought' in a moral way, it is by no means clear that we are rationally required to behave as we ought. Indeed, despite a great deal of effort, no philosopher has yet shown that the amoralist is necessarily irrational — much less that individual failures to be motivated by the demands of morality are necessarily irrational.<sup>33</sup>

The possibility that no special concern for morality is rationally required might, however, seem to form the basis for a new worry. This is the worry that if the non-cognitive attitude relevant to a particular evaluative term (here, 'immoral') is not rationally required, then we cannot expect the kind of overwhelming agreement that a response-dependent account presupposes. Suppose, for example, that moral requirements are not rational requirements. Then, even among rational people, there might not be any uniform response on which to base the ostensive teaching of moral terms. How then could we defend a response-dependent account of moral wrongness? Again, there are two responses. The first is again to concede that the proper account of a term like 'immoral' will not be a response-dependent one. But in this case, the concession need not entail that the proper account will therefore be an expressivist one. Rather, there is a third possibility, for which I have argued elsewhere: the proper analysis of 'immoral' might be one which involves a mix of both response-dependent evaluative terms, such as 'rational' and

---

33 See Sigrun Svavarsdóttir, 'Moral Cognitivism and Motivation,' *The Philosophical Review* 108 (1999): 161-219. For criticism of Christine Korsgaard's attempt to establish this strong result, see my 'Korsgaard's Private-Reasons Argument,' *Philosophy and Phenomenological Research* 64 (2002) 303-24. And there are other uses of 'ought' that are still farther removed from rational requirements: 'Bill ought to lose a couple of pounds,' 'You ought not wear that tie with that shirt.'

'harm,' and non-evaluative terms, such as 'equal concern.'<sup>34</sup> This would be an account of morality that had a response-dependent normative *basis*, but that did not offer a *direct* response-dependent analysis of moral terms. The second response to the new worry is to point out that even in the case of paradigmatic response-dependent terms such as 'green,' not all failures to respond in the characteristic way count as evidence of a visual defect. In learning color terms, we also learn what viewing conditions count as normal. If viewing conditions are not normal, then a sensation of blue in the presence of a green object need not indicate anything amiss. Similarly, it may be that for certain response-dependent evaluative terms, there are certain conditions that need to be met, beyond the mere rationality of the agent, before the agent's response would play the right kind of role in ostensive teaching. For example, when someone makes a wrong moral judgment because of anger, or because of special concern for one of the people involved, we can cite these distorting factors in order to dismiss the person's judgment. But neither anger nor special concern need be instances of *irrationality*. They can simply be sources of error. So even if not all rational people will respond uniformly to immoral action, it may still be true that all rational people who meet some further conditions respond with sufficient uniformity for the processes of ostensive teaching to yield a response-dependent referring term.

## VI Conclusion

The analogy between color and value has seemed useful to many philosophers. Both color and value seem obviously connected to standard human responses in some way. And yet color claims and evaluative claims seem capable of simple truth and falsity, independent of what may happen to human beings. Mars will be red after even our own stupidity, or nature's indifference, obliterates us from the face of the Earth. And even if we all became as corrupt and morally cynical as politicians, this would not make any difference to the fact that it is bad to be corrupt. One useful feature of a good analogy is that it gives one two places to look in the hopes of finding something clear about the structure or properties of the items that figure on each side of the analogy. I think that there is independent plausibility to the story I told about the way in which overwhelming agreement in response allows for the development of objective referring language. The plausibility of this

---

34 See my 'Problems for Moral Twin Earth Arguments,' *Synthese* 150 (2006) 171-83.

story, in the case of both color and value, supports the use of the analogy between color and value: the analogy is not simply brute, but has a foundation that can explain the limits of its usefulness. One of the useful features of the analogy is that it sheds light on the relation between having a certain response to an item, and uttering either a color ascription or an evaluative claim. The relation may, in each case, sometimes be that the utterance expresses the response. Yet, in the case of value the question of whether or not this relation obtains is obscured by the fact that there are more ways than one to determine whether or not something has any given evaluative property. The clarity of the relation in certain cases of color ascription is therefore helpful in showing that there is something important and correct in expressivism: in some cases, even though we may not be able to figure out which cases these are, the utterance of an evaluative claim may be the expression of a certain pro- or con-attitude. But this does not mean that such a claim would not also be the expression of a belief, any more than simple color identifications are not expressions of belief.

*Received: March 2005*

*Revised: September 2005*

