Non-Cognitivism, Higher-Order Attitudes, and Stevenson's "Do so as well!"

Meta-ethical non-cognitivism makes two claims - a negative one and a positive one. The negative claim is that moral utterances do not express beliefs thereby guaranteed to provide the truth-conditions for those utterances. The positive claim is that the primary function of such utterances is to express certain of the speaker's desire-like states of mind. Non-cognitivism is officially a theory of moral language, but non-cognitivists also maintain that moral judgements are themselves at least partially constituted by desire-like states to which moral utterances give voice. Non-cognitivists need a plausible account of what distinguishes whims, addictions and cravings from genuinely moral judgements. For while non-cognitivists maintain that in a suitably broad sense moral judgements just are constituted by desire-like states they also insist that not any old desire constitutes a genuinely moral judgement.[1] Since the challenge is to demarcate what is distinctive about the non-cognitive attitudes we might usefully call this the demarcation challenge.[2] One common strategy for meeting the demarcation challenge is to focus on higher-order desires and emotions – in particular, desires and emotions directed at getting others to share one's own desires and emotions. On its most plausible construal, the idea is that moral judgements necessarily involve higher-order attitudes, though further demarcations will need to be made if we are to be given necessary and sufficient conditions; for not all higher-order attitudes are moral judgements. This strategy has some of its earliest roots in Charles Stevenson's pioneering work. Stevenson argued that there is a 'do so as well!' aspect to moral discourse; that in saying something is good a speaker not only expresses

---

[1] Michael Smith has recently emphasized the importance of a plausible non-cognitivist way of meeting this challenge. See Smith 2002a and Blackburn 2002 for Blackburn's reply.
[2] James Lenman has usefully suggested this label for the problem; see his forthcoming reply to Smith 2002b.

his own first-order attitude of approval of the object of evaluation but also urges his interlocutors to share that attitude, thereby expressing a higher-order desire that they 'do so as well'.   More recently, in *Ruling Passions*, for example, Simon Blackburn emphasizes the importance of what he refers to as a "staircase of practical and emotional ascent." (Blackburn 1998: 9) Alan Gibbard explicitly refers to Stevenson's work and incorporates a 'do so as well!' element into his own theory (Gibbard 1990: 173) It is easy to see how one might plausibly think that this would help meet the demarcation challenge. After all, most appetites and mere whims do not, as a matter of empirical fact, satisfy this requirement. When I am hungry I typically do not desire that other people be hungry as well and when I form a fleeting desire to tap my pencil I do not typically desire that others desire that I do so or desire to tap their pencils too.

Another apparent attraction of meeting the demarcation challenge in this way is that also promises to help the non-cognitivist construct a theory of moral disagreement. The non-cognitivist needs an account of moral disagreement that extends to desires as well as beliefs. Dialectically, this issue is an important one because non-cognitivists often argue for their view on the grounds that it is uniquely situated to explain how moral disagreement can persist in spite of agreement on all of the relevant facts. One non-cognitivist strategy for understanding moral disagreement would be in terms of one person's desiring that someone else share her first-order desire. Once again, Stevenson's work picks up this theme. With a series of evocative examples, Stevenson argues that we do have a pre-theoretical and plausible notion of what he refers to as "disagreement in attitude." Moreover, Stevenson at one point suggests that two people disagree in attitude

when "at least one of them has a motive for altering or calling into question the attitude of the other." (Stevenson 1944: 3)

In spite of its attractions, this strategy for meeting the demarcation challenge is misguided. Stevenson's suggestion that there is a 'do so as well!' aspect to moral discourse is plausible enough, but does not help meet the demarcation challenge. For non-cognitivists should understand the desires involved in the Stevensonian 'do so as well!' not as constituting the speaker's moral judgement itself but as a pragmatic component of moral discourse. On Stevenson's account, to say something is good is to indicate one's approval of it and urge one's interlocutors to approve of it as well. In effect, Stevenson's insight is that in claiming that something is good a speaker expresses one's higher-order preference that others approve of it too.[3] It would be easy to move from Stevenson's insight about the function of moral language to the conclusion that to *think* something is good one must desire that others approve of it. However, this conclusion does not follow. The attitudes one must have to *think* x is good may not, in other words, be identical with the attitudes that one expresses whenever one *says* x is good. Moreover, anyone who holds that one must adopt such a higher-order attitude to count as thinking x is good faces a dilemma. Those higher-order attitudes are either instrumental or non-instrumental. If they are instrumental, then moral judgements so analyzed will often turn out to be irrational to accept in cases in which it seems manifestly obvious that they are not irrational. While if the relevant higher-order attitudes are non-instrumental, it would follow that to make moral judgements at all involves an implausible sort of fetishism. For on that account, to make a moral

---

[3] Such preferences are higher-order ones in that they are about other preferences, not in the sense that they are about the agent's own preferences.

judgement one must have a non-instrumental preference for something that one takes to have (at most) only instrumental value.   So the acceptance conditions of moral judgements do not include a higher-order desire that others share one's approval (or disapproval). The lesson for the non-cognitivist is that the non-cognitive attitudes constituting moral judgements may simply be certain kinds of first-order attitudes even if moral utterances conventionally express higher-order attitudes as well.

I.

It is plausible to suppose that one function of moral discourse is to persuade others to share one's non-cognitive attitudes, just as one function of descriptive discourse is to persuade others to share one's beliefs. In saying something is good, a speaker typically not only expresses her own approval of it, she also exerts pressure on her interlocutors to approve of it too. On Stevenson's account, a moral utterance expresses the speaker's non-cognitive attitude and expresses an imperative that her interlocutors share it. More recently, Alan Gibbard has picked up on this Stevensonian idea and built a similar element into his own version of non-cognitivism. Gibbard plausibly argues that in making a moral utterance part of what one is doing is, "demanding that the audience accept what he says, that it share the state of mind he expresses." (Gibbard 1990: 172) Gibbard glosses the speaker's demand that others share her attitude as an implicit claim of authority (Gibbard 1990: 174). I shall not argue for the Stevensonian view here, but simply grant it for the sake of argument. For my main concern is not to defend Stevenson's and Gibbard's positive claims about the function of moral language. Rather, my aim is to defend the negative thesis that we should not suppose that judging that something is good or bad essentially involves a pro-attitude in favor of others sharing

one's approval (or disapproval) of the thing being evaluated. I note the plausibility of Stevenson's and Gibbard's claims about the function of moral language only to emphasize that those claims, which are claims about the function of moral language, are consistent with the negative thesis about acceptance conditions defended here. Again, we should not exclude *ex ante* the possibility that while *saying* that X is good involves the expression of a higher-order attitude, *thinking* that X is good need not involve any such attitude. There are simply two different though related questions here: (1) What states of mind does a speaker express in saying that X is good (or bad)? (2) What makes it true that somebody thinks that X is good (or bad)?

We can accommodate the insight embodied in Stevenson's "Do so as well!" without building higher-order attitudes into our account of what it is to think that something is good. Here it is worth reviewing some of Paul Grice's rightly influential work in the philosophy of language. Grice argued that the states of mind a speaker expresses when making an utterance may not all figure in the truth-conditions of that utterance. If, for example, you ask me whether Russell Crowe is a good philosopher and I reply by saying, "Crowe has good handwriting," then it is clear enough that I have (in some sense) expressed the view that Crowe is not such a good philosopher. Nonetheless, the truth of my utterance does not depend on whether Crowe is a good philosopher. For what I literally said was that Crowe has good handwriting; whether that is true is simply a different question from whether he is a good philosopher. Grice drew this distinction in terms of the explicature of one's utterance - that which provides its truth-conditions - and that which is merely implicated (the "implicature"). More importantly for present purposes, Grice's examples also suggest a closely related but different point that Grice

did not explicitly discuss. For the examples show that we must distinguish between (a) the acceptance conditions for an utterance of a given type and (b) the states of mind expressed by a token utterance of that type. Presumably, for any "acceptance-apt" utterance- that is, for any utterance for which it is appropriate to talk of someone's accepting it - a person can be rightly said to accept such an utterance if and only if certain facts are true of her. I accept an utterance of 'grass is green' if and only if I believe that grass is green. Let us call the conditions individually necessary and jointly sufficient for someone to count as accepting an utterance (or judgement) the "acceptance conditions" for that utterance (or judgement). Grice's examples serve to illustrate that the states of mind you express with a given utterance may include states of mind one need not have to satisfy the utterance's acceptance conditions. For while I only need to think that Crowe has good handwriting to accept a token utterance of 'Crowe has good handwriting' my token utterance might well express a good deal more than my view of his handwriting.

Implicatures are usefully divided into conversational implicatures and conventional implicatures. The former depend heavily upon a context of utterance whereas the latter are a function of some fairly well-entrenched linguistic conventions surrounding particular linguistic devices and are not so context dependent. Words like 'but' and 'even' serve to introduce conventional implicatures. For example, an utterance of 'Even Crowe could make that inference' implies but does not literally say that Crowe is not very good at making inferences. One interesting feature of conventional implicatures is that unlike conversational implicatures they seem not to be cancelable. Whereas you can say, "Crowe has good handwriting, but I don't mean to suggest that he is not a good philosopher" without infelicity, it would be infelicitous to say, "Even Crowe

could make that inference, though I don't mean to imply that Crowe is not good at making inferences." The general point that the acceptance conditions for an utterance are distinct from the states of mind it expresses also holds with respect to utterances that are not truth-apt or at any rate do not express beliefs[4] that necessarily provide the utterances' truth conditions. So even if we are non-cognitivists about moral discourse we should distinguish the acceptance conditions for moral utterances from the states of mind they express. A failure to note this distinction can easily lead one to suppose that the acceptance conditions for moral utterances of the form 'X is good' include not only the approval of X but also the approval of people approving of X. This might seem tempting just because of the plausibility of the view that moral utterances typically express such higher-order attitudes – because of the plausibility of Stevenson's "Do so as well!"

Moreover, the very reasons typically given for supposing there is a "do so as well!" element to moral language provide us with equally good reason to resist the suggestion that thinking something is good essentially involves a "do so as well!" type desire. To make this point we must first mark an important distinction. We should preserve a distinction between judging that p, and judging that others should share one's judgement that p. In the case of ordinary empirical beliefs, this distinction is reasonable enough. Moreover, this point need not rest on the controversial supposition that there are no conceptual connections between thinking that p and taking oneself to be justified in thinking that p. For the point can be made even if we maintain the following conceptual

---

[4] In a suitably robust Humean sense of 'beliefs'; minimalists about truth and truth-aptness argue that any indicative sentence can serve to express a belief in a very thin sense. Non-cognitivism must be characterized in terms of beliefs in some richer sense. The standard move here is to invoke the idea of states with a particular "direction of fit" – beliefs are made "to fit the world" whereas desires are made "to make the world fit them." Cashing out the direction of fit metaphor is no easy task, though, and I shall not attempt to do so here. The arguments presented in the text should go through on any of a number of ways of drawing the distinction. For useful discussion, see Humberstone 1992 and Smith 1994.

connection between believing and justification: one can *come to* believe something only if one at the time at which one adopts the belief one takes oneself to be justified in so believing. This constraint should preserve as much of a connection between believing and taking oneself to be justified as is plausible. Fortunately, there are examples that respect this constraint but still make the point needed for purposes. For the constraint just articulated is temporally indexed to the time at which one forms a belief. The constraint therefore is compatible with beliefs sometimes having a sort of "inertia" once formed. It is consistent to think both that an agent must take herself to be justified in adopting a belief in the first place and yet hold that the belief can then take on a life of its own, outliving the agent's taking it to be justified. Indeed, this sort of phenomenon is not really controversial. For example, religious convictions that outlive their connection with the agent's epistemic commitments provide a fairly clear case of the phenomenon. Often the agent in such cases not only comes to believe there is no justification for her belief, in some cases she actually comes to think she has an all-things-considered justification for rejecting it. Still, the belief has a kind of inertia – she cannot, it seems to her at least, shake herself of it. Furthermore, that she continues to have this belief this seems borne out in her behavior (she continues to go to church, prays regularly, etc.). Nor does she seem to secretly believe that these beliefs really are justified in some way she cannot articulate – all of the available evidence indicates that she takes these beliefs to be unjustified. She not only will sincerely tell us that this is her view, we might add that she can give a detailed and persuasive account of the arguments against her view. This sort of thing does happen. One sometimes finds it in extended discussions of the philosophy of religion. Such people often distinguish faith from knowledge but typically do not

distinguish faith from belief. Indeed, they often emphasize that faith consists precisely in belief without epistemic justification, characterizing themselves as 'believers' as opposed to 'non-believers'. It is, I suppose, just possible that such people are deluded about their own states of mind, and do not really believe the things they say that believe, but this seems very implausible. Not only are people generally very reliable judges of their own beliefs, such people's behavior often seems best explained in terms of their actually believing what they say they believe – e.g., their going to church, praying, obeying certain otherwise pointless rules, etc. For present purposes, the point is that such a person need not suppose she has any authority on religious matters, that her religious belief is justified or that others should share her religious judgements.

These cases are relevant because we should make space for something analogous in the moral case. There are, after all, intuitive cases of this phenomenon in the moral domain too. I might no longer see any justification for thinking masturbation is immoral, for example, but find myself unable to shake loose of the judgement that it is morally bad. Early childhood conditioning sometimes leaves people in such a state. So it seems implausible to suppose that thinking something is good must be partially constituted by a desire that others approve of it. Why would anyone think otherwise? The justification most often suggested by the discussion in the literature is the idea that in judging that something is good one also implicitly sees oneself as an authority for that judgement, or at least supposes it is somehow justified and worthy of being adopted by others. The cases canvassed above in the case of ordinary beliefs should cast doubt on all of these theses – one can think that p without thinking that one's so thinking has authority, is worthy of being held or is justified. Crucially, this is perfectly compatible with accepting

the Stevensonian idea that there is a "Do so as well!" element to moral language. Generally speaking, if I say, "Abortion is bad," then I thereby express not only my disapproval of abortion but my desire that you share that disapproval. We rely on moral discourse to influence one another in this way, making various implicit conversational demands. This, however, is consistent with allowing that there is a logical gap between, e.g., judging that abortion is bad and judging that others ought to share that judgement. Allan Gibbard claims that in making a moral utterance part of what one is doing is, "demanding that the audience accept what he says, that it share the state of mind he expresses." (Gibbard 1990: 172) Once we distinguish this from Gibbard's more controversial claim that in making assertions we are claiming to be an authority of some kind, this idea is plausible, both with respect to the expression of ordinary descriptive beliefs and the expression of moral judgements. We can agree with Gibbard but hold that the higher-order attitude expressed in such a demand is not necessary for someone to count as thinking the object of evaluation is good.

In fact, the very reason Gibbard gives for including a "Do so as well!" element in his analysis of the function of moral utterances provides an equally good reason not to build such an element into our theory of their acceptance conditions. For if the "do so as well!" element is, as Gibbard argues, supposed to reflect the fact (simply assuming it is a fact) that one claims a kind of authority whenever one issues a moral utterance then the preceding discussion shows that this supposition of authority should not be built into the utterance's acceptance conditions. As we have seen, one can think that something is good (or bad) without seeing oneself as any kind of authority at all. Indeed, one can

think something is good and not even think one's view is justified. It is useful to return

to the non-moral case. Here is Gibbard:

> Conversation is full of implicit demands and pressures. Suppose I confidently expound astrology, and you give no credence. The result will be discomfort: in effect I demand that what I say be accepted, and you will not accede. (Gibbard 1990: 172)

The reason I only "in effect" make an "implicit" demand that what I say be accepted is

that this is a pragmatic element of what I say rather than a matter of its meaning. That

this is so is clear from the fact that my demand plays no part in the truth-conditions of my

utterance - there are possible worlds in which astrology is correct (let us suppose!) but in

which you do not give in to my demands, I do not make the demands, and I do not exist.

That this is not part of the acceptance conditions can also be seen from the way in which

it can without logical inconsistency be cancelled. "Astrology is an accurate way of

making predictions, but you shouldn't take my word for it, I'm irrationally biased on this

issue," is a logically consistent though odd thing to say. This sentence is vaguely similar

to the ones Moore discussed and that are now referred to as instances of "Moore's

paradox" – sentence of the form, "p, but I don't believe it" or "I believe that p but not-p."

In both cases there is something like a pragmatic tension or contradiction that falls short

of logical contradiction. Whereas, "Astrology is an accurate way of making predictions,

but its not" is logically contradictory precisely because it tries to cancel the expression of

a view that is essential to the acceptance conditions of the utterance. Similarly, in

making a moral utterance one typically makes an implicit demand that others share one's

attitude, but this is best understood as part of the pragmatics of your utterance rather than

its acceptance conditions. I might judge that abortion is bad but not think others should

share that judgement and therefore keep my mouth shut about it. Or if I say anything, I

might say, "Abortion is bad but don't believe it because its even worse to be the sort of person who has views on such matters." Again, this would be a very odd thing to say but it does not sound incoherent. Intuitively, the last clause of this utterance cancels any implication that one wants others to share one's judgement about abortion. It does raise the question of why the speaker said anything about abortion in the first place, but a suitable story can make sense of this (perhaps he was forced at gunpoint to divulge his considered opinion on the matter).

The expression of such higher-order attitudes is more analogous to conventional implicature than conversational implicature but differs from standard examples of conventional implicature in interesting ways. The "do so as well!" aspect of moral language is only analogous to these phenomena, strictly speaking, because the standard account of conversational and conventional implicatures is as expressing beliefs rather than desires. It is more analogous to conventional implicature because it does not depend much on the peculiarities of a particular context of utterance but is rather a function of the conventional meanings of moral terms when used in certain ways. This makes the "do so as well!" aspect of moral language an interesting case for the philosophy of language more generally. For it is usually supposed that conventional implicatures differ from conversational implicatures both in that they are not cancelable and in that they are less context-dependent. However, the "do so as well!" aspect of moral language is both conventional and cancelable. In this way, moral language seems to occupy a kind of Janus-faced relationship to standard ways of thinking about the differences between conventional and conversational implicatures. Like conventional implicatures, the "do so as well!" element is a function of the conventional meaning of some linguistic device in

the utterance in a way that is not highly context-sensitive, but like conversational implicatures the "do so as well!" is cancelable. Moreover, this feature of the "do so as well!" aspect of moral language should make it less controversial than standard cases of conventional implicature. For part of what makes those cases so controversial is the idea that though you cannot cancel the implicature it is still not part of what you literally said. This is controversial precisely because a defining mark of one's having literally said something simply is that one cannot cancel the implication without linguistic infelicity.[5] A nice feature of the sort of conventional implicature found in the moral case is that it avoids this problem insofar as it plausibly thought of as cancelable.[6] In any event, the main conclusion being defended here is that we can preserve the insights of Gibbard and Stevenson about moral practice by building the "Do so as well!" into the pragmatics of moral utterances rather than their acceptance conditions.[7]

II.

Let us take stock; so far two main points have been made. First, Stevenson's insight that moral language has an interesting 'Do so as well!' aspect does not warrant the conclusion that the acceptance conditions for a moral utterance include a higher-order preference that people share your first-order attitude. Any argument that takes Stevenson's insight as a premise and has the thesis that the acceptance of a moral utterance necessarily involves a higher-order attitude as its conclusion should be viewed with suspicion. So what seems like the only plausible source of an argument in favor of the view that accepting a moral utterance necessarily involves a higher-order attitude turns out to be unsound. Second,

---

[5] For a detailed case against the very idea of conventional implicature see Bach 1999.
[6] Note removed to preserve anonymity.
[7] Gibbard might agree, as he notes that conversational demands are "revocable" (Gibbard 1990: 172) in that a speaker can rescind those demands.

we have seen that the very reasons typically given for thinking that there is a "Do so as well!" element to moral language would provide equally good reason to resist the suggestion that thinking that something is good (or bad) necessarily involves any such higher-order preference. For example, suppose we agree with Gibbard that the higher-order attitude reflects one's taking oneself to be an authority of some kind (again, I do not mean to suggest that I do agree with this particular rendition of Stevenson's insight). In that case, building that higher-order attitude into the acceptance conditions would blur the distinction between thinking that p and thinking that one is an authority with respect to p. One might, of course, resist Gibbard's glossing of the Stevensonian idea as a claim of authority in the first place. A more modest and plausible gloss would be that the higher-order attitude reflects one's taking one's evaluation to be justified. However, that would also give us reason not to build such an attitude into our account of the utterance's acceptance conditions. For we have seen that there is possible to judge that p without supposing that one's judgement is justified. So the very reasons that might reasonably motivate a "Do so as well!" account of the function of moral utterances counsel against building such an element into our account of their acceptance conditions. In this section I argue that we should not, in any case, build such higher-order attitudes into our theory of acceptance conditions. Whereas in the previous section I offered a reply to what seems to be the only reason to include higher-order attitudes in our account of the acceptance conditions for moral judgements in this section I offer an independent argument that we should not include such attitudes.

A desire that others share one's approval (or disapproval) must be either instrumental or non-instrumental – that distinction is exhaustive. Of course, the non-

cognitivist might hold an ecumenical view according to which it does not matter whether the higher-order attitude is instrumental or non-instrumental; I consider this possibility below. Suppose the relevant higher-order preference is instrumental. To prefer something instrumentally is to prefer it as a means to something else. The question then arises, to what end is this preference instrumentally related? The only apparently principled answer would be that it is instrumental to whatever it is that the agent takes to be good. In that case, whenever I think that something, pleasure, say, is good I must also have an instrumental preference that others approve of pleasure, where I prefer their approval as a means to the promotion of pleasure. This, however, will mean that it will always be instrumentally irrational for me to think pleasure is good whenever I believe that others approving of pleasure will not promote (or even undermine) pleasure or something else that I think is good. To take an especially clear case, suppose that I am a hedonist and think that pleasure and only pleasure is good. Given the higher-order acceptance condition on moral judgement, it seems that I must thereby prefer that others approve of pleasure too and on this horn of our dilemma, this preference must be an instrumental one. Suppose, though, that I think that others approving of pleasure would undermine the promotion of pleasure in some way. Perhaps most people tend to be happier and to better promote overall happiness if they simply pursue their particular aims and projects while obeying certain moral rules than they would be if they actually aimed at pleasure itself. I might judge myself to be an exception to this general rule – while other people desiring pleasure is self-defeating my psychology is different in ways that mean this is not true of me. Suppose also that I know all of this. Whether such a scenario is at all empirically plausible is irrelevant; it is at least possible. In such a

scenario, it would be irrational for me to think that pleasure is good on the proffered interpretation of the higher-order attitude account of moral judgement. For in this situation to think pleasure is good is to prefer x (people approving of pleasure) as a means to y (pleasure) while believing that x will undermine y, and that is an instance of instrumental irrationality if anything is. This, however, is absurd. It may indeed sometimes be irrational to think that pleasure (or whatever) is good, but it should not be irrational just because one also thinks that other people preferring it will not advance pleasure (or any of one's other ends, for that matter).

This sort of objection can be pressed effectively regardless of the end to which one's preference is instrumental. Pick whatever end you like (call it E), and let the advocate of the higher-order acceptance conditions theory hold that judging that something (call it x) is good necessarily involves preferring that others share one's approval of E as a means to x. Then just take a case in which a person thinks x is good but does not believe that others approving of x would promote E, but rather thinks that their so approving would undermine the promotion of E. In such a case, it seems that any such agent is guilty of instrumental irrationality. For such a person should not be irrational just because they combine the judgement that x is good with the belief that other people's approval of x would thwart the promotion of some other end E. So the reductio of the proposed account of moral judgement easily generalizes. So we have good reason to reject the suggestion that value judgements necessarily involve a higher-order attitude. Or, rather, we have good reason to reject that account if the higher-order attitude is understood as an instrumental one.

Now consider the second horn of the dilemma. On this horn, the advocate of a higher-order attitude account maintains that to think that x is good one must have a *non-instrumental* preference that others share one's approval of x. On this sort of account, judging that x is good will not involve one in instrumental irrationality in cases in which it intuitively should not. So the advocate of a higher-order attitude account avoids the problems facing us on the first horn of our dilemma. On this account, though, it seems that moral judgements that seem perfectly reasonable actually involve one in a kind of fetishism. For it seems fetishistic to prefer something for its own sake when one judges it is good only as a means to something else. The miser, who desires money for its own sake even when he realizes (at some level) that it is good only as a means is an obvious albeit cliché case. The problem facing anyone who embraces the second horn of the dilemma I am pressing is that doing so seems to have the implausible consequence anyone who thinks that something is good is likely to be fetishistic. Again, suppose I think pleasure and only pleasure is good as an end. On the proposed account of value judgements, this means that I must have a non-instrumental preference that others approve of pleasure (though I may not care whether their approval is itself instrumental or non-instrumental). This, however, just is the sort of fetishism of which the miser is guilty. For if I think that pleasure and only pleasure is good then all else being equal I should prefer pleasure and only pleasure as an end.

There is a possible scope ambiguity here that must be avoided. The fetishism charge I am making is that it would be fetishistic for someone with certain values to desire-as-an-end that others approve of X. This must be distinguished from the rather different charge that it would be fetishistic for someone with certain values to desire that

others approve of X-as-an-end. In the first case, the fetishism attaches to my non-instrumental desire that others approve of X (regardless of whether their approval is instrumental or non-instrumental) whereas in the second case the fetishism charge attaches to my desire (which might be instrumental or non-instrumental) that others approve of X non-instrumentally. This is important to notice because on the second reading one might try to blunt the charge of fetishism by stipulating that when one thinks something is non-instrumentally good one must desire that others approve of it non-instrumentally but if one thinks that something is only instrumentally good that one need only desire that they approve of it, whether instrumentally or non-instrumentally. The main point here is that it is fetishistic for me to have a non-instrumental desire that others approve of X whenever it is only X that I take to have non-instrumental value and not others approving of X. In such a case, I may simply not give a damn whether others appreciate the value of X even if I think X really is good as an end.

A complication here is that such non-instrumental preferences for things the agent does not take to have non-instrumental value are not *always* implausibly fetishistic. For it makes sense for me to prefer something as an end that I do not take to be good as an end insofar as I believe that my so preferring will promote what I do take to be good as an end (again, pleasure, say). We should not miss the lesson of the so-called paradox of hedonism – the best way to promote one's own happiness almost certainly will involve the cultivation of non-instrumental preferences for things other than one's own happiness. In cases in which such preferences will advance one's own non-instrumental ends, the adoption of such non-instrumental preferences need not be implausibly fetishistic. Or, at any rate, if there is fetishism in such cases it is a fetishism that is rational to adopt, all

things considered. We can, however, control for this sort of phenomenon and still make the main point. For simply consider a case in which it is stipulated that a non-instrumental preference that others share one's approval of pleasure would itself undermine the promotion of pleasure, and the agent knows this to be true. In this case, to hold onto such a preference would be both fetishistic and irrational. This is another example in which intuitively the judgement that something is good need not carry with it any desire that others approve of it. However, if we accept the thesis that such a preference is partly constitutive of thinking pleasure is good, then it would follow that thinking pleasure is good necessarily would be fetishistic and irrational in such a case. Intuitively, this is absurd. There need be nothing fetishistic or irrational about thinking both that pleasure (and only pleasure) is good and at the same time thinking that your preferring that others also approve of pleasure would undermine the total amount of pleasure in the world. Hedonism presumably is sometimes self-defeating by the agent's lights in that the agent realizes that believing hedonism may undermine the promotion of hedonist ends. However, hedonism is clearly not self-defeating by the agent's lights *just* because the agent believes that her preferring that *others* approve of pleasure would reduce the amount of pleasure in the world. Intuitively that just gives her reason not to prefer that others prefer pleasure, and does not give her reason to abandon her acceptance of hedonism. Which is just to say that accepting hedonism intuitively is distinct from any desire that others share one's first-order attitude.

One could, of course, go for an ecumenical version of the higher-order attitude theory of acceptance conditions for moral judgement. On such an account, thinking something is good necessarily involves a preference that others approve of it, but it does

not matter whether the higher-order preference is instrumental or not. So long as one has such a preference one can count as judging that the object of evaluation is good, no matter whether the higher-order preference is instrumental. It should be clear that this ecumenical account cannot avoid the dilemma pressed here. For simply take those cases in which an agent reasonably and truly believes that others approving of something she values, beauty, say, would not advance any of her ends and indeed would thwart the promotion of beauty. If she thinks beauty is good, then given the ecumenical account she must nonetheless have some preference that others approve of beauty, whether that preference is instrumental or not. There will be a fact of the matter in each case – the agent's preference either will be instrumental or not. If the former, then theory is committed to positing instrumental irrationality where intuitively we should not. My thinking beauty is good as an end should not open me to a charge of instrumental irrationality just because I think that others approving of beauty would undermine the promotion of beauty and would not promote any of my other ends. However, if having an instrumental desire that others approve of beauty were part of the acceptance conditions for my judgement then I would be open to such a charge. If, however, the agent's desire is non-instrumental then the fetishism worry resurfaces; why should my positive evaluation of beauty as an end involve my approving of people approving of beauty as an end as well? Someone might well hold the view that beauty is good as an end *and* that others appreciating and approving of beauty is also good as an end – a kind of Moorean organic unity (see Moore 1903). This, however, need not be the view of the person we are imagining. Our hypothetical agent may well think that beauty and *only* beauty is good as an end. By her lights, the approval of beauty is good only insofar as it

serves to promote beauty. So it would be fetishistic for her to desire *as an end* that others approve of beauty. The point is not that the higher-order attitude view of acceptance conditions could not in any way distinguish these views. For on the higher-order attitude view, presumably an agent counts as thinking that the organic unity of beauty plus the approval of beauty by others is good must not only desire that others approve of beauty but also must desire that others approve of others approving of beauty. The point is rather that the higher-order attitude theory draws this distinction in the wrong place, bringing in either instrumental irrationality or fetishism where intuitively there should be neither. Either way, we have good reason to reject the higher-order acceptance theory of the acceptance conditions for moral judgements.

Conclusion.

Grice's discussion of implicatures shows that the acceptance conditions for a given utterance may not include all the states of mind such an utterance serves to express. Although Grice focused on truth-apt discourse, the distinction holds with respect to discourse that is arguably not truth-apt as well. In particular, we can accept the non-cognitivist thesis that moral language is not truth-apt and still distinguish the acceptance conditions for moral utterances from the states of mind expressed by such judgements. In the moral case this is just to say we should distinguish, e.g., what states of mind one gives voice to in saying that something is good and what states of mind one must have in order to count as thinking that thing is good. A failure to heed this distinction can lead one to suppose that the plausible Stevensonian thesis that there is a "do so as well!" element to moral discourse entails the implausible consequence that to think that something is good (or bad) one must prefer that others approve (or disapprove) of it. Fortunately, this

consequence follows only if we fail to internalize Grice's insight and its applicability to forms of discourse that are not truth-apt. Once we see that this insight also applies to moral discourse even if that discourse is not truth-apt we can reject the thesis that moral utterances have higher-order attitudes built into their acceptance conditions. Non-cognitivists can happily accept the Stevensonian thesis that there is a "do so as well!" element to moral language without building a higher-order desire that others "do so as well" into the acceptance conditions of moral utterances. Since such a theory of the acceptance conditions is independently implausible for reasons canvassed here, this good news for non-cognitivists. However, this also means that the non-cognitivist cannot explain what is distinctive about the attitudes that constitute moral judgements themselves in terms of higher-order attitudes. Nor, given this line of argument, should the non-cognitivist adopt the otherwise tempting strategy of explaining moral disagreement in terms of conflicting higher-order attitudes. For we can have moral disagreements that remain unvoiced. Indeed, if I think that something is bad and someone else thinks it is not bad then we disagree even if we have never heard of one another, much less discussed the issue. So non-cognitivists cannot fully explain moral disagreement in terms of speech-acts and the higher-order attitudes they express. This is not necessarily such bad news for non-cognitivists. In my view, non-cognitivists can more plausibly meet the demarcation and provide a plausible account of moral disagreement without invoking higher-order attitudes. That, however, is another story.

Bibliography

Bach, K. 1999. "The Myth of Conventional Implicature." *Linguistics and Philosophy*. 22. 327-366.

Blackburn, S. 1998. *Ruling Passions*. Oxford: Clarendon Press.

_____. 2002. "Replies." *Philosophy and Phenomenological Research*. 65.

Gibbard, A. 1990. *Wise Choices, Apt Feelings*. Oxford: Clarendon Press.

Grice, P. 1989. *Studies in the Ways of Words*. Harvard University Press: Cambridge, MA.

Humberstone, L. 1992. "Direction of Fit." *Mind*. 59-83.

Moore, G.E. 1903. *Principia Ethica*. New York: Cambridge University Press.

Smith, M. 1994. *The Moral Problem*. Oxford University Press: Oxford.
_____. 2002a. "Which Passions Rule?" *Philosophy and Phenomenological Research*.
        65. 157-163.
_____. 2002b. "Evaluation, Uncertainty, and Motivation." *Ethical Theory and Moral
        Practice*. 5. 305-320.
Stevenson, C. L. 1944. *Ethics and Language*. Yale University Press: New Haven, CT.