European Journal of Philosophy

DOI: 10.1111/j.1468-0378.2011.00455.x

Hybrid Expressivism and the Analogy between Pejoratives and Moral Language

Ryan J. Hay

Abstract: In recent literature supporting a hybrid view between metaethical cognitivism and noncognitivist expressivism, much has been made of an analogy between moral terms and pejoratives. The analogy is based on the plausible idea that pejorative slurs are used to express both a descriptive belief and a negative attitude. The analogy looks promising insofar as it encourages the kinds of features we should want from a hybrid expressivist view for moral language. But the analogy between moral terms and pejorative slurs is also problematic. In this paper, I argue for two main ways in which we should distinguish between two different types of pejorative terms: slurs, on the one hand, and what I call general pejorative terms, on the other. I examine the problems with the analogy between slurs and moral terms and conclude that general pejorative terms like 'jerk' are a better candidate on which to model the potential dual-use behavior of moral terms. So if hybrid theorists are looking for a dual-use model for moral language, they should be careful to base their analogies on general pejoratives, rather than slurs.

I Introduction

A natural observation about racial and ethnic slurs is that they are often used by speakers to both describe their targets and express negative attitudes. For instance, when a speaker utters the sentence 'Leonardo is a wop', she is neither solely describing Leonardo as Italian nor solely expressing an attitude akin to that which she would express with the sentence 'Boo Leonardo!' So the term 'wop' is neither purely descriptive nor purely expressive. When a bigoted speaker utters such a sentence about Leonardo, she both describes him as Italian and expresses some sort of negative attitude about Italians. This point has been recently noticed and exploited by those who think that moral language has similar dually descriptive and attitudinally expressive features. The analogy between pejoratives and moral language initially looks promising, but I argue that we must be careful to distinguish between different types of pejorative terms. While much of the current literature on hybrid theories has focused on racial and ethnic slurs, I argue that terms like 'jerk' and 'asshole'—terms which I

call *general pejoratives*—are a better model for a *dual-use* account of moral language.

This paper overviews the philosophical attractions of such 'dual-use' or 'hybrid' metaethical theories, which benefit from an analogy between moral language and pejoratives. I argue for two main ways in which we should distinguish between slurs and general pejorative terms. With regard to both of these distinctions, moral terms work like general pejoratives, rather than like slurs. So if hybrid theorists are looking for a dual-use model for moral language, they should be careful to base their analogies on general pejoratives, rather than slurs.

A. Cognitivism, Expressivism, and Hybrid Views

Historically, there has been a clear divide between cognitivist and expressivist metaethical positions. As Michael Ridge writes: 'Cognitivists insist that moral utterances express beliefs rather than desires, while expressivists hold that moral utterances express desires rather than beliefs' (Ridge 2006: 304).[1] On an expressivist picture of moral semantics, the state of mind a sentence is conventionally used to express tells us something fundamental about that sentence's *meaning*. Given this type of picture, expressivists provide a radically different account of the type of meaning possessed by moral language in contrast to ordinary descriptive language.

There are familiar advantages and challenges faced by each type of position. Cognitivists are able to account for why simple moral statements appear to be descriptive, and furthermore face no particular challenge with the logic of moral language since they hold that moral terms work just like ordinary descriptive terms. It seems plausible to think that one is describing an action when one labels it 'wrong' or 'right', and it seems natural to think that one has *beliefs* about what is wrong or right in an analogous way to how one has beliefs about what is green or square. So cognitivist positions have both theoretical advantages and intuitive appeal. But such positions have a difficult time accounting for the apparent emotional and motivating dimensions of moral language, and some have gone so far as to deny that it even has these features.

Noncognitivist expressivists, on the other hand, are primarily motivated by the idea that moral language has these features, and their theories attempt to explain how this could work. And it also seems plausible, after all, to think that calling something 'right' or 'wrong' is to express some sort of pro or con attitude about it. So expressivist positions also look like they have both theoretical advantages and intuitive appeal. But expressivists face the additional explanatory burden of addressing what has come to be widely known as the *Frege-Geach problem*: they must make sense of moral predicates as they appear in complex unasserted contexts, such as in constructions involving standard logical operators. For example, while expressivists hold that a sentence like 'stealing is wrong' has a similar expressive function to a sentence like 'Boo stealing!', it is

much more difficult for expressivists to explain the expressive function—and hence the meaning—of sentences like 'if stealing is wrong, then I shouldn't do it', or even 'stealing is not wrong'. Expressivists such as Simon Blackburn and Allan Gibbard both have proposed solutions to the problem,[2] but their respective solutions are neither straightforward nor unproblematic.[3]

Over the course of the past decade, a range of views have been proposed that attempt to incorporate the theoretical advantages and intuitive appeal of both cognitivist and noncognitivist positions, while avoiding the pitfalls associated with each. Many of these views hold that there is at least some noncognitive element (such as a motivational state, desire-like attitude, or a commendatory or condemning function) associated with moral language in addition to whatever standard descriptive features it possesses. The specifics of such *hybrid views* vary widely, but all such views attempt to draw from the resources available to both expressivist and ordinary cognitivist positions.

David Copp has been one of the main advocates for such a view. In both his earlier and more recent work on the subject (2001, 2009), Copp argues that the noncognitive component of moral discourse is expressed via conventional implicature. Similarly, Stephen Barker (2000) has proposed an implicature theory for moral language on which sincere utterances of sentences like 'stealing is wrong' explicitly express that stealing has a certain property, and conversationally implicate that the speaker is committed to the disapproval of things that have that property. Michael Ridge (2006) endorses a view on which moral sentences *express* (in the expressivist's semantic sense of 'express') both beliefs and desires. Ridge claims his 'ecumenically expressivist' view is able to solve the Frege-Geach problem 'on the cheap' if we accept a certain notion of validity (one that does not necessarily involve valid arguments being truth preserving). Hybrid expressivist views have been proposed elsewhere in recent metaethical literature (Alm 2000, Boisvert 2008). In addition, Stephen Finlay (2004, 2005) holds a cognitivist view on which there is attitudinal content associated with moral language that gets communicated via conversational implicature. Dorit Bar-On and Matthew Chrisman's Neo-Expressivist view (2009) is fundamentally cognitivist about moral semantics, but maintains that speakers (non-semantically) express relevant motivational states by the act of making moral utterances.

B. The Analogy between Moral and Pejorative Language

Several proponents of hybrid views have explicitly claimed that moral terms work in an analogous way to pejoratives. For instance, Copp writes that 'typical moral predicates, such as ''morally right'', are similar to pejorative terms in that their use can both ascribe a property and express a relevant conative attitude' (Copp 2009: 170). Copp tells us that pejoratives such as 'wop'[4] 'illustrate the basic idea' of a hybrid view. The basic idea is that *dual-use expressive terms* (a class that from the point of view of the hybrid theorist includes both moral terms and pejoratives), possess a *descriptive component*, which ascribes a property in just the

same way ordinary descriptive predicates do (e.g., '... is green'), and an *attitudinally expressive component*, which is associated with a relevant non-cognitive attitude. Boisvert, in defending his own hybrid view, writes that:

> It is uncontroversial, for example, that [literal utterances] of sentences containing emotionally charged predicates, such as sentences used to spout racial epithets (e.g., 'Bob is a ____'), *both directly describe certain people as having a certain property* (e.g., as being of a certain race or ethnic group), and *directly express the speaker's contempt toward anyone that has that property* (and not just toward the subject of the sentence). (Boisvert 2008: 180)[5]

So, to call someone a 'wop' is both to ascribe to them the property of being Italian, and to express a negative attitude about Italians. The hybrid hypothesis is that similarly, to call an action 'wrong' (or 'right') is to ascribe to it a certain property, and to express a negative (or positive) attitude about actions which have that property.

If this type of account is on the right track, then hybrid views that endorse the analogy have a straightforward response to the Frege-Geach problem. Boisvert has noted that pejorative terms like 'wop' express the same negative attitudes even when they occur embedded in more complex constructions (e.g., 'If Bob is a wop, I'd be surprised'). Pejoratives such as 'wop' express a negative attitude about anything that has the property it is used to pick out, and such attitudes are expressed even when the term is embedded in complex unasserted contexts. There is no special problem explaining the meanings of complex sentences containing pejoratives, so if moral terms work in the same way, then there is no special problem for explaining the compositionality of their meaning, either.

But Mark Schroeder has recently questioned the limits of the analogy between moral and pejorative language made by hybrid theorists (Schroeder 2009, 2010). He has argued that while the apparent dual-content nature of pejoratives may offer what he calls a 'license for optimism' for the hybrid theorist, there are relevant differences between the dual-use features of pejoratives and the dual-use features hybrid theorists should want moral language to have. If so, pejorative and moral language may be fundamentally different enough that we should call into question the recent attempts to model moral terms on how slurs work. Part of my task in this paper, then, is to examine the particular problems that arise in the analogy between moral and pejorative language. I say more about this in both Sections II and III.

But the analogy is far from completely doomed. I maintain that there are important distinctions to be made between different types of pejorative terms. Much of the hybrid theorists' discussion of pejorative terms has focused on racial and ethnic slurs. In contrast to these, I maintain that there are *general pejoratives terms*, for instance terms like 'jerk', 'asshole', 'nerd', 'loser', 'dummy', and 'tool', that share relevant features with moral terms, and that exhibit precisely the same characteristic differences from slurs. In contrast to slurs, we *can* make suitable

analogies between the dual-use function of moral terms and general pejoratives. So another goal of this paper is to make clear the relevant distinctions between slurs and general pejoratives. Sections II and III are primarily devoted to drawing two distinctions between slurs, on the one hand, and moral terms and general pejoratives, on the other. Section II distinguishes slurs from general pejoratives by examining the nature of their descriptive components. Section III distinguishes slurs from general pejoratives by examining how their respective attitudinally expressive components embed in belief reports. Plausible hybrid accounts of moral language should recognize that with regard to both distinctions, moral terms work like general pejoratives. In Section IV, I examine the potential relations between these two distinctions.

II Descriptive Detachability

A. Slurs and General Pejoratives

While all pejorative predicates plausibly look to be dual-use in the sense of being both descriptive and attitudinally expressive, different pejoratives exhibit characteristic differences when it comes to identifying their descriptive components. Let us begin by considering the descriptive component of the sentence *Wp*:

Wp: 'Pavarotti is a wop.'

In this case, there is clearly a sentence involving a neutral predicate (that is to say a predicate that does not express any noncognitive attitude) that has precisely the same extension as the original sentence's descriptive component. This sentence is just *Ip*:

Ip: 'Pavarotti is Italian.'

Mark Richard makes the point that: 'Every slur, so far as I can tell, has or could have a ''neutral counterpart'' which co-classifies but is free of the slur's evaluative dimension' (Richard 2008: 28). *Ip* is the *neutral counterpart sentence* to the sentence *Wp*. The neutral counterpart of any sentence involving a given evaluative predicate is obtained by replacing the evaluative predicate in question with a new predicate involving a term that is coextensive with the descriptive component of the original evaluative term, and that lacks any evaluative dimension.

With regard to racial slurs, Jennifer Hornsby writes: 'It is as if someone who used [a slur] had made a particular gesture while uttering the word's neutral counterpart. An aspect of the word's meaning is to be thought of as if it were communicated by means of this (posited) gesture' (Hornsby 2001). It will be useful here to adopt Blackburn's (1992) convention of using the down arrow symbol [↓] to indicate a further negative tone, inflection, or gesture on an

occasion of utterance. Blackburn introduces the notation as an addition to the descriptive term 'fat':

> I shall transcribe 'fat' said with a sneer as 'fat ↓,' where the downward arrow signals the combination of emphasis on the first consonant and downward cadence that carries the sneer. 'fat ↓ ' will be heard most often in the mouths of those who are repelled by or despise fatness ... (Blackburn 1992: 290)

Hornsby's observation amounts to the claim that a term like 'wop' would have the same kind of meaning as 'Italian ↓ '. In addition, ordinary descriptive terms can be accompanied, on their occasions of use, by a tone, inflection, or gesture in order to communicate a further negative attitude. While I grant that such things are not really terms in the sense that 'wop' or 'jerk' is a term, I will refer to these as 'arrow terms' for convenience. Arrow terms seem to be relevantly similar to slurs in that they have obvious neutral counterparts. Their neutral counterparts are just the ordinary descriptive term without the accompanying tone, inflection, or gesture.

Now consider the descriptive component of *Jp*:

> *Jp*: 'Pavarotti is a jerk.'

What is the sentence that bears the neutral counterpart relation to *Jp*? The answer here proves to be much more difficult than in the former case involving 'wop'. I maintain that there is no neutral counterpart sentence for 'Pavarotti is a jerk' that involves any *preexisting* term coextensive with the descriptive component of 'jerk'. For this reason, I say that the descriptive component of 'jerk' is *not detachable*. I use the term 'detachable' here in a special sense, without regard to any previous technical usage, as follows:

> *Descriptive Detachability*: The descriptive component of a dual-use evaluative term T is detachable if and only if there is some preexisting neutral counterpart coextensive with T.

By this definition, racial and ethnic slurs like 'wop' possess detachable descriptive components, while *general pejorative terms* like 'jerk' and 'asshole' do not.

The claim that the descriptive component of 'jerk' is not detachable warrants further justification, but I want first to make two brief points here regarding the descriptive component of general pejoratives. First, even though I hold that general pejoratives do not have detachable descriptive components, they can have varying degrees of descriptive specificity. For example, 'nerd' appears to be more narrowly focused than 'jerk'. Second, while I hold that there is no *preexisting* neutral counterpart term for 'jerk', it does not follow from this that no neutral counterpart term could be given. Under the assumption that a given term is really dual-use, we can always introduce a new term into the language that refers to the description associated with a given dual-use evaluative. Using David Kaplan's (1978) 'dthat' operator, the neutral counterpart sentence for

'Pavarotti is a jerk' is just 'Pavarotti instantiates dthat[the property expressed by the descriptive component of 'jerk']'. But such a sentence is not very informative as to what the descriptive content of 'jerk' is, and cannot escape reference to the original pejorative term. Mark Richard makes a similar point. In reference to slurs possessing neutral counterparts, Richard maintains that something similar is true for terms like 'asshole' (Richard 2008: 28). Richard does mention that it might be an interesting exercise to spell out what the descriptive content associated with 'asshole' is, but avoids attempting to perform such an exercise himself (Richard 2008: 34). My method provides us with a simple way to provide a neutral counterpart for general pejoratives like 'jerk' and 'asshole', but it is also compatible with the thesis that there is no *pre-existing* neutral counterpart term.

Now, let us return to consider the specific claim that 'jerk' does not have a detachable descriptive component. Perhaps there are those for whom this does not seem obvious. Suppose that one could provide a neutral counterpart sentence for *Jp* with a sentence of the form:

Γp: 'Pavarotti is a Γ'

where Γ is some already existing descriptive term. Let us consider several potential candidates for Γ. To begin with, there are terms that might be roughly synonymous with 'jerk', take 'asshole' for example, but these terms are clearly attitudinally expressive in just the same way 'jerk' is (if not even more negatively so in the case of 'asshole'). The candidates for Γ must be neutral descriptive terms.

So Γ could be something like: 'liar', or 'self-centered', or 'unreliable', or 'hyper-aggressive', etc. But unless one has an unjustifiably narrow conception of precisely what a jerk is, it doesn't look like any single one of the above could be correctly taken to be the neutral counterpart for 'jerk'. Clearly someone who is self-centered, or unreliable, or hyper aggressive may very well be a jerk, but any one of these descriptions alone is not descriptively equivalent to 'jerk' in the way Italian is descriptively equivalent to 'wop'.

So maybe Γ cannot be replaced by a simple descriptive term, but a complex disjunction of some set of preexisting descriptive terms. For example, maybe we can get a complex neutral counterpart sentence for *Jp* by giving some long sentence like 'Pavarotti is [self-centered OR unreliable OR hyper aggressive OR (loud AND boastful) etc.]' This might seem promising, but there are two problems with it. First, figuring out precisely what set of terms should be disjoined is no easy task. The task is made difficult by the fact that the relevant descriptions associated with the use of the pejorative term 'jerk' can seem to vary depending on the context in which the term is used. Should the list of descriptive terms comprising the set associated with the use of 'jerk' be composed of every possible description associated with every possible circumstance of 'jerk's normal use as a pejorative predicate? If so, the disjunction would be very long, and this leads us to the second problem. If the neutral counterpart of 'jerk' is given by some relatively long disjunction—for example: (A OR B OR C OR D OR E)—then to say 'Pavarotti is not a jerk' ($\sim Jp$) is to say something quite strong: it is to say

Pavarotti is [$\sim$ A and $\sim$ B and $\sim$ C and $\sim$ D and $\sim$ E]. But this just doesn't look like what speakers assert with $\sim Jp$. So a complex disjunctive view looks implausible for the descriptive content of 'jerk'.[6]

Perhaps Γ could be the description for which the term was originally coined. In the case of 'jerk', the origin of the term is unclear, but this is not the case for other similar pejorative terms. Consider the term 'geek'. In this case, the term was originally used to refer to carnival performers who would bite the heads off of chickens. So the neutral counterpart sentence for 'Pavarotti is a geek' might be 'Pavarotti is a chicken-head-biter-offer'. But this is clearly *not* the description *now* associated with *normal* uses of the pejorative predicate '... is a geek'. Many pejorative terms that are currently employed in the English language have nothing to do with whatever description they were originally associated with upon their introduction into the language. Kaplan makes a similar point about the use of the pejorative term 'bastard'. He points out that while the description 'one born out of wedlock' may 'figure etymologically in the genesis' of the pejorative use of the term, it does not correctly account for what it currently means as a pejorative term (Kaplan 2004: 5).

All of this might lead the reader to the conclusion that the difference between slurs like 'wop' and pejoratives like 'jerk' is that the former are dual-use while the latter are not. Perhaps there is no neutral counterpart to 'jerk' because it lacks a descriptive component, and we should be pure expressivists about such general pejorative terms. Aside from the theoretical Frege-Geach style problems such a view would face, it seems intuitively implausible. Surely there is *some* descriptive content associated with the use of general pejorative terms like 'jerk'. Consider, for a moment, 'jerk' in contrast to a range of other general pejorative terms; such as 'asshole', 'bastard', 'nerd', and 'geek'. I think it is clear that each of these terms has a negative attitude associated with their normal use. But is that all they have? If these terms are *only* used to express negative attitudes and nothing more, then it would seem as if the only difference between these pejorative terms could be the degree to which such a negative attitude is expressed by their use. Or perhaps the expressivist about pejoratives could maintain that different pejoratives express different negative noncognitive attitudes. But neither of these non-cognitive interpretations of general pejorative terms is satisfactory. One need go no further than the local school yard to have the specific descriptive differences between jerks, nerds, and geeks explained. Such terms clearly have some descriptive content, in addition to a negative attitude, associated with their use as pejorative terms. The question is: how do we understand specifically what the descriptive content is? This is an easy task for slurs like 'wop', but remains much more difficult for general pejoratives.

B. Moral Terms

Moral terms like 'wrong' are similar to general pejorative terms like 'jerk', rather than slurs like 'wop', with respect to descriptive detachability. Consider an

example of a sentence involving a moral term, for instance: 'Stealing is wrong.' Is there a different neutral sentence, involving a descriptive term that has precisely the same extension as the descriptive component of 'wrong'? The answer to this question is at least as complicated and problematic as it was for 'jerk'. Moral terms like 'wrong' do not have detachable descriptive components.

We can consider precisely the same kind of responses here that I originally considered in relation to general pejorative terms. The idea that 'wrong' has a preexisting neutral counterpart is not plausible for reasons analogous to those that led us to reject the corresponding claim about 'jerk'. Perhaps those who agree with certain simple and strict consequentialist or deontological accounts for the meaning of moral terms[7] would object, but such views are unjustifiably narrow in just the same way as a view that holds that the descriptive component of jerk is just coextensive with some particular preexisting descriptive term. Furthermore, the descriptive component associated with the use of 'wrong' should not be analyzed with an eye toward its specific etymological development as a term within the English language, nor does its descriptive essence seem easily captured by some long disjunction of descriptions.

Copp appears to recognize that there is some difference between slurs and moral terms specifically with regard to what I call the detachability of their descriptive components. Copp makes a similar point using the terminology from his own view: 'One might wonder why there is no familiar predicate that stands to ''wrong'' as 'Italian' stands to ''wop''. Why is there no familiar term with the same core meaning as ''wrong'' but without its simplicature meaning?' (Copp 2009: 188). He attributes this concern to Kent Bach in a footnote, but does not spend very much time addressing it in his paper. Copp quickly concludes that: 'Of course, even though there is no familiar term that is a neutral equivalent for ''wrong'', there are neutral equivalent phrases, such as ''has the property called 'wrongness'''' (Copp 2009: 188). Notice that Copp's thought here appears to be in line with my view that sentences involving moral terms have neutral counterparts, but that there is no pre-existing coextensive term that is the neutral counterpart for a moral term (like 'wrong'). This being said, Copp drops the point soon after raising it, and never discusses the properties of 'jerk' in relation to this concern. Perhaps if he had, he would have realized that 'jerk' seems to exhibit an analogous character to 'wrong' in this regard. But Copp's question remains: why do some evaluative terms have descriptive components that are detachable while others do not? In the next sub-section, I will turn to briefly consider a potential answer to this question before returning (in Section III) to discuss another important distinction between slurs and general pejorative terms.

C. *A Potential Underlying Explanation for Descriptive Detachability*

Many may be inclined to think that the issue of descriptive detachability is somehow related to what we might call the *variability* of the description associated with a given evaluative term. The idea here is that unlike other types

of evaluative terms, slurs have a very *fixed* description associated with their use. In contrast, one might think that there is some greater or lesser degree of variability in the descriptions associated with terms like 'jerk' and 'wrong'. This would explain the ease with which we can identify a pre-existing term coextensive with the descriptive content of 'wop' (i.e. Italian), while one cannot do the same for a term like 'jerk'; in the latter case there simply isn't a single description picked out by 'jerk'. And one might plausibly suppose that the description associated with 'jerk' varies depending on the context in which the term is used. If the difference in the detachability of a dual-use predicate's descriptive component is just a matter of this kind of variability, then perhaps this difference between slurs and general pejoratives is not so significant.

After all, there are purely descriptive terms that work in an analogous fashion. Consider the descriptive term 'large'. There is a certain variability in the description picked out by the use of 'large' depending on the context in which claims of the form: 'x is large' are uttered. For instance, consider claims like 'the *USS Enterprise* is large' when discussing naval vessels versus the claim 'Omicron Ceti is large' when discussing stars. Clearly, the standards a star has to meet to qualify as large are different from the standards that must be met in order for a naval vessel to be large, even though 'large' is used in both instances to pick out an object of relatively greater size than the other objects in its class. But even if the class of objects is fixed by the context, there is still a certain vagueness associated with terms like 'large'. For instance, suppose one were to pose the question: 'What is the largest US naval vessel?' One could imagine an answer of the form: 'Well, the *USS Enterprise* is longer and wider, but the *USS Ronald Reagan* is heavier'. And such an answer could plausibly be followed with the question: *'What exactly do you mean by "large"?'* If this kind of view for large seems plausible, then it is also plausible to think that the description 'large' is variable in multiple senses: it can vary based on context, and it can possess a certain amount vagueness about what it picks out.

One might be inclined to think that 'jerk' works in a similar fashion, but in addition is used to express a negative attitude. For instance, consider a claim like 'that driver is a jerk' versus claims like 'don't date Pavarotti, he is a jerk.' Presumably, the standards an individual has to meet to qualify as a jerk in the context of motoring are different from the standards an individual has to meet in the context of relationships, even though the use of 'jerk' to classify an individual in both cases pertains to the qualities of the individual and a negative attitude about those qualities. And, just like the case involving variable purely descriptive terms, 'jerk' may also admit of a certain degree of vagueness about the qualities it picks out. We could similarly imagine a conversation like the following:

Speaker A: 'Is Jack or Pavarotti a bigger jerk?'
Speaker B: 'Well, Jack is generally inconsiderate, but Pavarotti is always loud and obnoxious; I guess *it depends on what you mean by "jerk"*.'

So, on this view, the descriptive content of 'jerk' is variable in just the way a purely descriptive term like 'large' could have a variable content.

The advantage of such a view is that an inherent variability in the descriptive content of certain evaluative terms may be able to account for their character with regard to descriptive detachability. In cases where an evaluative term's descriptive content is variable, it seems plausible to assume that there would be no obvious neutral counterpart term. So, it would be just those terms with variable descriptive contents that would fail the test for descriptive detachability.

But while this might appear to provide a deeper explanation for why some terms fail the descriptive detachability test, descriptive variability and descriptive detachability can in fact come apart. Let us more closely examine the assumption that descriptive variability could underlie descriptive detachability. If this were the case, then evaluative terms would fail the test for descriptive detachability as a result of the variability in their descriptive contents.

But there are counterexamples to this type of claim. Consider arrow terms involving terms which exhibit descriptive variability. For instance, we could imagine something like 'large ↓' (which picks out the same thing as 'large' and in addition expresses a negative attitude). In this case, the descriptive content of 'large ↓' is variable insofar as the content of the purely descriptive term 'large' is. But clearly 'large ↓' passes the test for descriptive detachability; it possesses a descriptive component that is coextensive with a preexisting descriptive term—namely 'large'. So, if we consider things like 'large ↓' to be at least significantly like other dual-use evaluative terms, then the mere variability of such a term's descriptive content is not alone enough to result in it failing the test for detachability.

So the descriptive detachability of a given dual-use predicate cannot be explained solely by appealing to descriptive variability. But descriptive detachability is still an independent and important way to distinguish between different dual-use evaluative terms. I will now turn to discuss what I take to be another significant way to distinguish between dual-use terms.

III Belief Ascription Embedding

A. Slurs and General Pejoratives

In Section II we saw that moral terms are much more like general pejoratives than like slurs, with respect to the ease with which we can find a term for its neutral counterpart—what I've called their 'detachability'. I will now turn to argue that moral terms are also much more like general pejoratives than like slurs, with respect to the way that they behave in belief reports. I will argue that belief reports involving arrow terms and slurs can be used to express the negative attitudes of just *the speaker* of the sentence, while in contrast belief reports involving general pejoratives and moral terms appear to attribute negative attitudes to their subjects. I will proceed by making several observations about belief reports containing the types of evaluative terms we have been discussing up to this point.

To begin with, consider the following examples of belief reports involving an arrow term and corresponding neutral description, as well as a slur and its neutral counterpart:

B($L \downarrow p$):	'Jack believes that Pavarotti is large $\downarrow$.'
B(Lp):	'Jack believes that Pavarotti is large.'
B*(Wp)*:	'Jack believes that Pavarotti is a wop.'
B*(Ip)*:	'Jack believes that Pavarotti is Italian.'

Consider a speaker who utters the sentence B($L \downarrow p$). The use of a sneer or gesture by the speaker is entirely compatible with the speaker expressing her own negative attitudes, and reporting Jack's neutral beliefs about Pavarotti's relative size. In this kind of case, notice that B($L \downarrow p$) *can be used to express the speaker's* negative attitude, rather than to attribute any negative attitude to Jack. It seems especially clear in this case that the speaker can express her own attitude by uttering B($L \downarrow p$) rather than B(Lp) to report Jack's belief. After all, it is the speaker who is sneering or gesturing while making the utterance.

The case with the slur term 'wop' is relevantly parallel to the previous example. Suppose that Jack believes that Pavarotti is Italian, and Jack has no bigoted attitudes about Italians. You and I would be inclined to use the neutral *B(Ip)* to characterize Jack's belief. But it is plausible to think that a bigot would report Jack's belief by uttering *B(Wp)*. After all, a bigot may be inclined to use 'wop' instead of 'Italian'. So the bigot can use *B(Wp)* to express his own bigoted attitudes, rather than attribute a bigoted attitude to Jack. If a speaker's utterance of *B(Wp)* always attributes a negative attitude to Jack, then the bigot's use of *B(Wp)* would be incorrect; Jack has no bigoted attitudes about Italians. But the bigoted speakers' use of B(*Wp*) *is entirely compatible* with Jack's belief that Pavarotti is Italian.

This gives us reason to think that *B(Wp)* can be used in such a way that it both (1) expresses the negative attitudes of its speaker, and (2) does not attribute any negative attitude to its subject (Jack). Furthermore, if we change the scenario so that Jack is a bigot, and perfectly willing to go around stating things like 'Pavarotti is a wop', those who do not share his attitudes toward Italians would be unwilling to characterize Jack's beliefs by stating the sentence *B(Wp)*, in order to avoid being seen as expressing negative attitudes about Italians. Perhaps the point is even clearer if you pause to consider whether or not you would, in any conversation, ever use *B(Wp)* to characterize the beliefs of Jack the anti-Italian bigot, without any further statements qualifying your use of the slur.

In contrast to the cases involving slurs and arrow terms, consider a case involving the term 'jerk' embedded under 'believes that':

B*(Jp)*: 'Jack believes that Pavarotti is a jerk.'

Suppose that Jack believes that Pavarotti has a certain property. Suppose that it is just the property that jerks have which makes them jerks. Let us also suppose that

Jack is just a really nice guy and has no negative attitude about anyone who possesses that property. In fact Jack only has a positive attitude about everyone, regardless of the properties they possess. Unlike the previous case involving slurs, it looks as though *B(Jp)* could never be appropriately uttered. It is not plausible to think that we (those who have negative attitudes about jerks) would ever characterize Jack's belief by uttering *B(Jp)*, even though we all would call anyone who has that certain property a 'jerk'. One would simply never use the sentence *B(Jp)* to *only* express one's own negative attitude about jerks. In fact, Jack doesn't think anyone is a jerk, even though he is able to recognize that some individuals possess properties which are specifically the ones you and I pick out by the use of the term 'jerk'. This gives us reason to think that *B(Jp)* is only used in such a way that it always attributes a negative attitude to its subject. If the speaker utters the sentence *B(Jp)*, the speaker is attributing to Jack not only the belief that Pavarotti has a certain property, but also the negative attitude about individuals who have that property.[8] This is to say that a speaker's use of *B(Jp)* attributes to the subject of the sentence both the belief in the descriptive component, as well as the desire-like attitude 'jerk' is used to express.

In the next sub-section, we will see that both Copp and Boisvert think that moral terms embedded under 'believes that' work in a similar way to how I maintain that general pejorative terms like 'jerk' work. We will also see that Schroeder (2009) has argued that this is precisely the kind of belief embedding behavior that hybrid theorists should hold moral terms exhibit, if they want to be entitled to several of the explanatory advantages that hybrid views are thought to have.

B. *Moral Terms, Belief Embedding, and the Big Hypothesis*

The conclusions of the previous section present a challenge for those wishing to support hybrid analyses for moral terms based on an analogy with slurs. Similar considerations appear to worry both Copp and Boisvert, and Schroeder explicitly presses hybrid theorists about belief ascription embedding. For instance, Boisvert concludes that: '"John believes that donating to charity is right,'' is true if and only if the *subject* of the sentence is in a psychological state that is the same in content as the ascription's compliment sentence' (Boisvert 2008: 195).[9] The generalized form of such a claim is what Schroeder labels the 'Big Hypothesis' (Schroeder 2009: 301):

> Big Hypothesis: If 'P' is a sentence expressing mental states $M_1 \ldots M_n$, then the descriptive content of 'S believes that P' is that the referent of 'S' is in each of mental states $M_1 \ldots M_n$.

Schroeder explains that what the Big Hypothesis tells us 'is that attitude ascriptions (at least, ''believes that'' ascriptions) attribute not only belief in the descriptive content of the complement clause, but the desire-like attitudes that it expresses, as well' (Schroeder 2009: 301).

For instance, on a hybrid view for moral terms that endorses the Big Hypothesis, the sentence 'Max believes that stealing is wrong' attributes to Max the view that stealing has a certain property (which, for the sake of convenience, we can label *K*), and also attributes *to Max* a negative attitude about actions that have that property. In contrast, belief reports involving slur predicates can fail to obey the Big Hypothesis, since slurs can be used by speakers to express their own negative attitudes even when reporting the neutral beliefs of others. Schroeder recognizes this, though he makes no explicit distinction between slurs and general pejoratives. In reference to dual-use theories such as Copp's and Boisvert's that are in some sense modeled on how slur predicates are supposed to work, Schroeder writes:

> if 'wrong' really works like a pejorative, then we should not expect it to obey the Big Hypothesis, either. This is bad for the hybrid theory if it is true. . . . it fails to make good on the Big Hypothesis, and hence on the key potential attractions for the hybrid theory. (Schroeder 2009: 305)

I agree with Schroeder that 'wrong' doesn't work like a *slur*. Instead, I hold that moral terms work like general pejoratives; both 'wrong' and 'jerk' conform to the Big Hypothesis about dual-use evaluative terms embedded under belief ascriptions, while sentences involving slur predicates embedded in belief reports can clearly fail to do so.

Why should we value a hybrid model in which moral terms conform to the Big Hypothesis? According to Schroeder, there are two main advantages that hybrid views have if they endorse the Big Hypothesis about belief reports. First, it can be appealed to in order to explain certain aspects of open question phenomena. Schroeder points this out with the following examples (Schroeder 2009: 301):

1 Max believes that stealing is *K*, but does Max believe that stealing is wrong?
2 Max believes that stealing is wrong, but does Max believe that stealing is wrong?
3 Max believes that stealing is wrong, but does Max believe that stealing is *K*?

Notice that these examples involve a version of the open question phenomenon involving 'believes that' ascriptions. On a hybrid view for moral terms that endorses the Big Hypothesis about belief ascriptions, we can translate these sentences as follows (Schroeder 2009: 301):

1′ Max believes that stealing is *K*, but does Max believe that stealing is *K* and desire not to do what is *K*?
2′ Max believes that stealing is *K* and desires not to do what is *K*, but does Max believe that stealing is *K* and desire not to do what is *K*?
3′ Max believes that stealing is *K* and desires not to do what is *K*, but does Max believe that stealing is *K*?

With regard to this set of examples translated using a hybrid view which conforms to the Big Hypothesis, Schroeder writes that:

> the second question 'answers itself' in a way that the former question does not—even once we are clear-headed about the descriptive content of 'wrong'. The same reasoning also predicts the intuitive asymmetry between 1 and 3, because unlike 1, sentence 3 does 'answer itself'... (Schroeder 2009: 301)

So it appears that a hybrid view which endorses the Big Hypothesis gives us a unique way to explain away the appearance of at least some types of open question phenomena; at the very least, those involving belief reports.

The second main advantage of a hybrid view that endorses the Big Hypothesis is that it presents a straightforward explanation for motivational internalism. Given that moral terms obey the Big Hypothesis, anyone who satisfies 'x believes that stealing is wrong' will also satisfy 'x believes that stealing has a certain property, and x has a negative desire-like attitude about anything that has that property.' So in this case, x already has the relevant belief and desire-like attitude to motivate her not to steal, assuming that there is a connection between an agent's beliefs and attitudes and what motivates an agent to perform (or not perform) an action.[10] So a hybrid view for moral terms which endorses the Big Hypothesis possesses advantages in both explaining away certain types of open question phenomena and providing a straightforward explanation of motivational internalism. Insofar as hybrid theorists want to take advantage of these explanations, this presents a challenge for those wishing to support hybrid views for moral language based on an analogy with slurs.

In the next subsection, we'll see that both Boisvert and Copp note the unique behavior of slurs embedded under belief ascriptions. Boisvert attempts to explain why slurs, specifically, may fail to obey the Big Hypothesis. Copp also recognizes that terms like 'wop' behave characteristically different from terms like 'wrong' and 'jerk' when embedded under 'believes that'. I shall now turn to these points.

C. *Boisvert's Wide-Scoping Explanation*

Boisvert does recognize and attempt to explain away the worry that belief ascriptions involving slur predicates don't seem to attribute negative attitudes to the subject, but rather express the speaker's attitudes. He asks us to consider the following examples involving slurs and belief attributions (Boisvert 2008: 193)[11]:

B52. Jackie believes that wops eat a lot of pasta.

B54. Jackie believes that Jesus-freaks are correct in their beliefs.

Again, notice that a *speaker* who utters sentences like B52 or B54 appears to be the one expressing the negative attitude in relation to the relevant group. According to Boisvert, 'These examples provide some evidence for the claim that a

complement sentence's expressive component is being expressed by the speaker rather than attributed by ''believes that'' to the subject of the sentence' (Boisvert 2008: 194). Boisvert accounts for this by arguing that there is a scope ambiguity involved in the use of slurs embedded under 'believes that'. For instance, in the case of B52, Boisvert claims that the quantified noun phrase 'wops' takes wide scope relative to 'believes that'. He holds that we interpret B52 as B52* (Boisvert 2008: 194):

B52*. [All x: x is wop][Jackie believes that x eats a lot of pasta].

Boisvert generalizes based on these examples and concludes that 'the appearance that complement sentences containing a quantified noun phrase are used to express a speaker's conative state is easily explained as taking wide-scope relative to ''believes that''(Boisvert 2008: 194).

Unfortunately, however, there are counterexamples to Boisvert's scope ambiguity claim. There are examples of 'believes that' attributions involving slurs where the quantified noun phrase takes narrow scope relative to 'believes that' but the speaker who utters it is still the one expressing the negative attitude. Consider the following sentence as uttered by a speaker:

J. Jackie believes that Christians are going to heaven.

Does the quantified noun phrase 'Christians' take wide or narrow scope relative to 'believes that'? Here are the two possibilities:

J_{Narrow}. Jackie believes that [all x: x is a Christian][x is going to heaven]
J_{Wide}. [All x: x is Christian][Jackie believes that x is going to heaven]

I contend that the narrow scope reading is more natural here. Presumably, a speaker uttering J is not trying to say something about the beliefs Jackie has about *each* person who in fact is a Christian. Suppose that Jackie mistakenly believes that Barack Obama is a Muslim, and furthermore believes that Muslims are not going to heaven. In this case, the wide scope reading of J is false; Jackie does not believe that Barack Obama is going to heaven even though Obama is in fact a Christian. However, the narrow scope reading of J appears to be true if Jackie does believe that Christians are going to heaven. Now, if we replace 'Christian' in J with the pejorative slur 'Jesus-freak', then we get the sentence:

J*. Jackie believes that Jesus-freaks are going to heaven.

A speaker who has a negative attitude about Christians may use sentence J* to report Jackie's belief that Christians are going to heaven. The appropriate account of J* holds that the quantified noun phrase 'Jesus-freaks' takes narrow scope relative to 'believes that' for precisely the same reason that the quantified noun phrase 'Christians' takes narrow scope relative to 'believes that' in sentence J. This is an example where a quantified noun phrase may take narrow scope relative to 'believes that' but still be used to express the *attitude of the speaker*,

rather than the attitude of the subject of the complement clause. So the explanation of why slurs can fail to attribute negative attitudes to the subjects of belief ascriptions cannot rest solely on the idea that slur-associated quantified noun phrases take wide scope over 'believes that'.

Copp also realizes that slurs appear to behave differently from other types of pejoratives in the way they embed under 'believes that'. Copp writes: 'I may simplicate[12] that I have contempt for Italians in saying ''Bill thinks you are a wop'' but I do not simplicate that I have contempt for anyone in saying ''Bill thinks you are a jerk''(Copp 2009: 186–7). Copp's comments here are strikingly similar to the difference he points out between how 'wop' and 'wrong' embed under belief ascriptions. He also points out that 'a speaker who says ''Betty believes Leonardo is a wop'' simplicates that she [the speaker] has contempt for Italians but a speaker who says ''Anna believes capital punishment is wrong'' does not simplicate that she [the speaker] disapproves of capital punishment' (Copp 2009: 187). But Copp does not continue on to draw any fundamental parallel between the semantic usage rule he thinks is associated with 'jerk' and that associated with 'wrong' based on their similarity in belief ascription embedding. Instead he sketches slightly different semantic usage rules for each. But on my view, such a parallel between 'wrong' and 'jerk' is unsurprising, and points to a greater semantic similarity between 'wrong' and 'jerk' than either one has with 'wop'.

So, if a hybrid theorist is looking to draw parallels between moral terms and other dual-use evaluative terms, general pejorative terms like 'jerk' are much better candidates than slurs. 'Wrong' works more similarly to 'jerk', both in the way that they embed under belief ascriptions, and, as we saw in the previous section, with respect to the non-detachability of their descriptive components. In the next section, I turn to examine a potential relation between descriptive detachability and belief embedding behavior.

IV The Relation between Descriptive Detachability and Belief Embedding

In the preceding two sections, I have argued that there are two important distinctions to be made between different types of dual-use evaluative terms. Up to this point, I have said nothing about how these distinctions might be related. But this is important. If these distinctions are unrelated, then we should expect to find examples of dual-use terms which have detachable descriptive components and always attribute attitudes to the subjects of belief reports, or examples of dual-use terms which do not have detachable descriptive components but can be used to express the attitudes of just the speaker even when the speaker utters the term in a belief report. But there are no clear examples of either type. All of the classes of potentially dual-use terms we have examined up to this point either have detachable descriptive components and can be used to express the attitudes of the speaker when embedded in belief ascriptions (arrow terms and slurs), or fail the test for descriptive detachability and conform to Schroeder's Big

Hypothesis about belief ascriptions (general pejoratives and moral terms). So this makes it seem as though these distinctions must be related.

One way in which they might be related is that one distinction could be used to explain the other. There is a line of thought (present, as we shall see, in the current hybrid literature) that could give us a way to explain belief embedding behavior by appealing to descriptive detachability. This involves a certain plausible hypothesis about what *speakers* can communicate by *choosing* certain terms over others. This point, taken in conjunction with the idea that the descriptive detachability of a particular evaluative term is connected to the *ready availability* of a neutral counterpart term, could explain why terms with detachable descriptive components can be used to express the attitudes of speakers even when embedded in belief ascriptions. In this section, I will examine how such an explanation could proceed. However, given another assumption that I will argue is at least highly plausible, the explanation is problematic. So, in the end, I point out a limitation on the relation between these two distinctions, and maintain that more research needs to be done in order to address fully the questions of whether and how they are related.

A. The Speaker's Choice Hypothesis

The basic idea about how to explain belief embedding in terms of descriptive detachability proceeds as follows: If a dual-use evaluative term is descriptively detachable, then that term has a preexisting neutral counterpart term that is extensionally equivalent but has no attitudinally expressive element to its use. And if a dual-use term has a preexisting neutral counterpart, then there is a *choice* between the neutral and non-neutral term when picking out the same descriptive content. So, for instance, a speaker has a choice between uttering *B(Wp)* and *B(Ip)* when reporting Jack's beliefs about Pavarotti. Since the speaker has this *choice*, it is reasonable to think that the speaker would specifically choose to use *B(Wp)* when she (the speaker) wants to express a negative attitude about Italians. The key idea is that if a speaker can *choose* between an evaluative term and a readily available neutral counterpart, then the speaker's use of the evaluative construction results in the expression (via conventional implicature) of the attitude associated with the term by the speaker. So, if a dual-use evaluative term is descriptively detachable, then the speaker's use of the evaluative construction results in the expression of the attitude associated with the term by the speaker.

Stephen Finlay points out that a speaker's choice between a slur term and a neutral counterpart term can explain the idea that the negative attitude associated with a slur term gets expressed by conventional implicature. Finlay writes:

> Consider again pejoratives, a conventional-implicature account of which I acknowledge to be not without merit. If pejoratives do indeed carry colouring conventionally, it is partly because they exist in the language as alternatives to other words with the same denotations. Why would a

> speaker call a person a 'faggot' rather than a homosexual, or a 'nigger' rather than a Black or African-American? This choice of terminology is explained by the intention to express contempt towards a group. (Finlay 2005: 19)

I take it that this comment is meant to apply generally to any cases where the speaker has a choice between the use of a slur-term and a neutral counterpart term. Sentences involving slurs embedded under 'believes that' would be simply more specific instances in which a speaker may choose between a slur and a neutral descriptive term. But Finlay's general claim could apply just as well in these cases. If Finlay's view is correct here, then it could also be used to explain why a sentence like *B(Wp)* would express the speaker's contempt for Italians; after all, the speaker could have chosen to talk about Jack's belief about Pavarotti's national origin by using *B(Ip)*.

Whether or not a conventional implicature account should be taken as the correct view for the conveyance of whatever noncognitive element is associated with dual-use terms is still an open question; but such a view does appear to have the advantage of being able to explain why a *speaker's choice* to use an expressively charged slur term rather than a neutral descriptive term results in the expression (via conventional implicature) of the speaker's attitude.

In contrast, there is no analogous choice that can be made between sentences involving evaluative predicates that do not possess detachable descriptive components and their corresponding neutral counterpart sentences. Consider B(Jp) and B(Γp):

> *B(Γp):* 'Jack believes that Pavarotti is a Γ.'

As we saw in Section II, figuring out exactly what the neutral content is to 'jerk' (Γ) is problematic in a way that it simply isn't for 'wop'. Since there is no readily available neutral counterpart term, there is no readily available neutral counterpart sentence for the speaker to use instead of the non-neutral evaluative in order to pick out the right content. So, what we now have is a specific sort of hypothesis about the relation between descriptive detachability and belief-ascription embedding: we can explain the latter by appeal to the former. Let us call this the *Speaker's Choice Hypothesis* (SCH):

> SCH: In cases where a speaker can *choose* between an evaluative term and a readily available neutral counterpart, the speaker's use of the evaluative construction results in the expression (via conventional implicature) of the attitude associated with the term by the speaker.

While SCH looks plausible, there are several issues that arise with this hypothesis upon further consideration. When SCH is considered in conjunction with the idea that there may be expressive variability between different general pejoratives that are descriptively equivalent (for instance 'jerk' and 'asshole'), it looks like

something similar to SCH would result in confusion about the expressive content of general pejoratives.

B. *Speaker's Choice and Expressive Intensity*

It is plausible to think that there are some sets of general pejorative terms that can vary from one another only by the intensity of the attitudes that they are associated with, rather than any variation in their descriptive content. For instance, 'asshole' is in some sense a more extreme pejorative term than 'jerk', and the sense in which it is more extreme could be due to the attitudes each term is associated with. Furthermore, it is conceivable that there is no descriptive difference between the terms 'jerk' and 'asshole', in much the same way there is no descriptive difference between 'Italian' and 'wop'. The difference between these two pairs, however, is that while 'Italian' is neutral and 'wop' is additionally attitudinally expressive, both 'jerk' and 'asshole' are attitudinally expressive. The difference between the general pejorative terms could be, then, a difference in degree between the strength of the attitudes each terms is used to express, respectively.

Now such an assumption might not strike the reader as very intuitive; after all, while one might think all assholes are jerks, one might be less inclined to think that all jerks are assholes. But such an intuition does not mean that there is an extensional difference in the descriptive content of 'jerk' and 'asshole'. One may not be inclined to apply the term 'asshole' in just the same instances one is inclined to use the term 'jerk', but this has to do with the attitudes of the speaker, rather than the extension of the descriptive content associated with each term. Analogously, one may not be inclined to use the term 'wop' in just the same cases where one is inclined to use the term 'Italian', but this likewise is due to the attitudes of the speaker, rather than the extension of the descriptive content of the terms.

If belief ascription embedding could be explained by the availability of a choice on the part of the speaker in the way supposed by SCH, there would also be a similar choice that could be made between sets of general pejoratives that differ only in the intensity of the attitudes associated with their use. Consider the following sentence involving the general pejorative term 'asshole':

B(Ap): 'Jack believes Pavarotti is an asshole.'

Now, on the assumption that 'asshole' is an expressively stronger way of saying 'jerk', the speaker has a choice between *B(Ap)* and *B(Jp)* when attributing the belief to Jack. Following the same line of reasoning used in the previous cases, one might think that a speaker would choose *B(Ap)* as opposed to *B(Jp)* to express her own (the speaker's) negative attitude about the properties picked out by general pejorative terms like 'jerk', 'asshole', and 'fucker'. So the *stronger* negative attitude associated with 'asshole' would get attributed to the speaker, since the

speaker could have chosen the sentence *B(Jp)* rather than *B(Ap)* to talk about Jack's belief.

But this analogous account is potentially problematic. Based on the argument I made in Section III, the negative attitudes associated with general pejorative terms get attributed to the subject of the sentence. But if something like the SCH is a correct account of why the speaker is the one expressing the attitude in the case of slurs embedded under belief ascriptions, then it looks like similar reasoning could lead one to hold that it is the speaker that expresses a stronger negative attitude when using *B(Ap)* instead of *B(Jp)*. So, in any case where there are sets of general pejorative terms that differ from one another in terms of the intensity of the attitude they are used to express, the speaker has a choice to use a term with less expressive intensity when uttering the belief attribution. In the case where a speaker uses *B(Ap)* to talk about Jack's belief that Pavarotti' is a jerk, it looks as if the speaker would *both* be attributing a weaker negative attitude to Jack, as well as expressing a stronger negative attitude herself.

If that is right, then making out precisely the attitudes that get attributed versus expressed starts to become exceedingly difficult. For instance, would it be the case that there is some minimal attitude that is always attributed to the subject of the belief attribution sentence involving general pejorative terms? Furthermore, if this is the right account, then the proposed relation between belief ascription embedding and descriptive detachability, as it was proposed in SCH, is no longer plausible. Obviously, even if 'jerk' and 'asshole' pick out the same descriptive contents, they still fail the test for descriptive detachability.

Conclusion

In this paper, I have maintained that pejoratives are both descriptive and attitudinally expressive. Since hybrid theorists think something similar holds for moral language, it has been natural for them to look at pejoratives as a model for how dual-content language works. But I have argued that we should distinguish between different types of pejoratives. What I call general pejorative terms—like 'jerk'—have descriptive components that are not detachable, and, when embedded in belief reporting sentences, the negative attitudes they are used to express get attributed to the subjects of such sentences. In contrast, slurs (and arrow terms) have detachable descriptive components, and they can be used by speakers to express *their own* negative attitudes even when reporting the beliefs of others. On the hybrid assumption that moral language really is dual-content, then with regard to both of these distinctions it functions like general pejorative language, rather than slurs or arrow terms. So hybrid theorists should move away from recent attempts to model the dual-content nature of moral language on slurs like 'wop', and look instead to general pejoratives, like 'jerk'.

For dual-content terms, the connection between descriptive detachability and belief embedding remains unsettled. I have examined one plausible explanation: that the presence of a choice between a neutral descriptive term and its

non-neutral dual-use counterpart determined whether or not the speaker expresses the attitudinal content when making belief reports. But, as we saw, if this reasoning were correct, then it looks like similar considerations would apply to belief ascriptions involving more and less expressive general pejorative terms. But such considerations conflict with our intuitions about how we use general pejoratives in belief reports.

Finally, the proposed connection discussed in Section II-C between descriptive variability and descriptive detachability does not provide us with any account of the belief ascription embedding behavior of evaluative terms. If 'jerk' has a similar kind of descriptive variability to an arrow term like 'large ↓', then this phenomenon applies to terms that can exhibit different belief ascription embedding behaviors. Therefore descriptive variability cannot, at least, by itself, be an underlying explanation for either descriptive detachability or belief ascription embedding behavior. But even though the relationship between these phenomena remains unexplained, it is still clear that these distinctions are important, and advocates of hybrid theories would do well to pay attention to them.[13]

Ryan J. Hay
School of Philosophy
University of Southern California
USA
ryanhay@usc.edu

NOTES

[1] My emphasis on 'express' in both instances

[2] See Blackburn 1984, Gibbard 1990.

[3] See van Roojen 1996; on the negation problem for Gibbard's moral semantics see Unwin 2001.

[4] The term 'wop' or 'WOP' is an ethnic slur for Italians or persons of Italian heritage. According to the Urban Dictionary ⟨http://www.urbandictionary.com/define.php?term=wop⟩, 'It is often said that this term comes from ''With Out Papers'' or ''Working On Pavement''', but the Merriam-Webster dictionary provides the following account of its etymology: 'Italian dial. *guappo* swaggerer, tough, from Spanish *guapo*, probably from Middle French dial. *vape, wape* weak, insipid, from Latin *vappa* wine gone flat'⟨http://www.merriam-webster.com/dictionary/wop⟩.

[5] My emphasis on 'both directly describe certain people as having a certain property', and 'directly express the speaker's contempt toward anyone that has that property'.

[6] I do not wish to maintain that straightforwardly reductive or disjunctively reductive analyses of the descriptive component are the only acceptable options here, though these are the main candidates I discuss. In general, this discussion is not meant to be exhaustive of the potential attempts to provide a neutral counterpart for a general pejorative term like 'jerk'.

[7] For instance, those who think that *the property of wrongness* is just coextensive with the property of failing to maximize happiness or not following from a universalizable maxim, respectively . . .

[8] Some may be inclined to think that when a speaker utters *Jp*, she is expressing a negative attitude just about Pavarotti, rather than individuals who have, for lack of a better term, 'jerky' properties. This alternative position is precisely the type of view that David Copp appears to endorse. Copp initially attributes the idea that 'jerk' might work differently from slurs to Kent Bach in a footnote:

> Kent Bach suggested to me in conversation that it might be useful to distinguish between two kinds of pejorative terms. There are (a) terms, such as 'Yankee' . . . that are used to refer contemptuously to a class of persons or things such that their use typically expresses or implies contempt for all persons or things in that class; and (b) terms, such as 'jerk,' that are used to refer contemptuously to persons or things such that their use implies that the speaker has contempt for the person or thing explicitly referred to, but does not imply that she has contempt for anyone or an (Copp 2001: 19 fn)

If Copp's early thought on the distinction between slur terms like 'Yankee' and pejorative terms like 'jerk' is correct, then it would appear as if moral terms actually work more like slurs than general pejoratives like 'jerk'. Copp is maintaining here that pejorative terms like 'jerk' do not express attitudes about anyone who has a certain property, but rather a specific individual.

But, if this idea is correct, we can simply reconstruct the embedding problem just for dual-use evaluatives that function in this way. For instance, consider the argument:

P1. If Bill is a jerk, then Pavarotti is a jerk.	[Jb → Jp]
P2. Bill is a jerk.	[Jb]
C. Pavarotti is a jerk.	[Jp]

Specifically, one could not easily explain why anyone who accepts the premises should be forced to accept the conclusion. For if one accepts the conclusion, then, on the dual-use view for 'jerk' in question, one would have to accept its associated descriptive component and have a negative attitude toward Pavarotti. While one could easily explain why anyone who accepts P1 and P2 must also accept the descriptive component associated with C, it is completely unclear how the negative attitudes associated with P1 (presumably, no negative attitude toward anyone) and P2 (negative attitude toward Bill) would lead one to have a negative attitude toward Pavarotti. So any account of the predicate '. . . is a jerk' which treats it as expressing contempt only toward its subject must bear the additional burden of having to provide a special account for how it embeds in more complex sentences.

This being said, I think Copp's view here attempts to account for the intuitively plausible assumption that a speaker uses a sentence like *Jp* to indicate a negative attitude toward Pavarotti. But my claims here about the dual-use nature of 'jerk' can account for this as well. Let us return to the original supposition I made in Section III-A: that speakers use *Jp* to express a belief that Pavarotti has a certain property (or set of properties) and a negative attitude about anyone who has those properties. If this is correct, then consequentially the speaker does indicate contempt for Pavarotti—since the speaker both believes that Pavarotti has a certain property and condemns (or has some negative desire-like attitude about) anyone who has it. In just the same way, the speaker who utters *Wp* consequentially indicates a negative attitude about Pavarotti.

[9] My emphasis on 'subject'.

[10] This is directly following the line Schroeder (2009) takes on pp. 302–3. On p. 303, Schroeder writes that under the assumption 'that the nature of beliefs and desires is to motivate us, other things being equal, to do what, given the truth of the beliefs, would attain the object of the desire' and in conjunction with a hybrid view which endorses the Big Hypothesis, we get a strong form of motivational internalism, namely, the claim that: 'Necessarily, for all x, if x believes that stealing is wrong, then x will be motivated, other things equal, not to steal.'

[11] I follow the numbering convention from Boisvert's original paper here.

[12] For Copp, 'conventional simplicature' is a modification to Grice's notion of conventional implicature which takes into account some of Kent Bach's challenges from 'The Myth of Conventional Implicature' (1999). Copp writes: 'A speaker *conventionally simplicates* that p in assertorically uttering a sentence just in case *(a)* in assertorically uttering the sentence, the speaker communicates that p (whether intentionally or not), and *(b)* the fact that the speaker thereby communicates the proposition p is determined by the (or a) conventional meaning of some particular linguistic device in the sentence, but *(c)* the proposition that p is not part of what is said explicitly by the speaker in uttering the sentence, so that *(d)* the falsity of p is compatible with the truth of what is said explicitly by the speaker' (Copp 2009: 184).

[13] I am particularly grateful to Mark Schroeder for his input and feedback on this paper. I would also like to thank Stephen Finlay, Sam Shpall, Geoff Georgi, Indrek Reiland, and the anonymous referees from the *European Journal of Philosophy* for their comments.

REFERENCES

Alm, D. (2000), 'Moral Conditionals, Noncognitivism, and Meaning', *Southern Journal of Philosophy*, 38: 355–77.

Bach, K. (1999), 'The Myth of Conventional Implicature', *Linguistics and Philosophy*, 22: 327–66.

Barker, S. (2000), 'Is Value Content a Component of Conventional Implicature', *Analysis*, 60: 268–79.

Bar-On, D. and Chrisman, M. (2009), 'Ethical Neo-Expressivism', in R. Shafer-Landau (ed.) *Oxford Studies in Metaethics: Volume 4*. New York: Oxford University Press.

Blackburn, S. (1984), *Spreading the Word*. Oxford: Oxford University Press.

—— (1992), 'Morality Through Thick and Thin', *Proceedings of the Aristotelian Society*, 66: 285–99.

Boisvert, D. (2008), 'Expressive-Assertivism', *Pacific Philosophical Quarterly*, 89: 169–203.

Copp, D. (2001), 'Realist-Expressivism: A Neglected Option for Moral Realism', *Social Philosophy and Policy*, 18: 1–43.

—— (2009), 'Realist-Expressivism and Conventional Implicature', in R. Schafer-Landau (ed.) *Oxford Studies in Metaethics: Volume 4*. New York: Oxford University Press.

Finlay, S. (2004), 'The Conversational Practicality of Value Judgment', *The Journal of Ethics*, 8: 205–23.

—— (2005), 'Value and Implicature', *Philosopher's Imprint*, 5: 1–20.

Gibbard, A. (1990), *Wise Choices, Apt Feelings*. Cambridge: Harvard University Press.

Hornsby, J. (2001), 'Meaning and Uselessness: How To Think about Derogatory Words', *Midwest Studies in Philosophy*, 25: 128–41.

Kaplan, D. (1978), 'Dthat', in P. Cole (ed.) *Syntax and Semantics, Volume 9: Pragmatics*. New York: Academic Press.
—— (2004), 'What is Meaning? Explorations in the Theory of Meaning as Use', Unpublished Manuscript.
Richard, M. (2008), *When Truth Gives Out*. New York: Oxford University Press.
Ridge, M. (2006), 'Ecumenical Expressivism: Finessing Frege', *Ethics*, 116: 302–36.
Schroeder, M. (2008), *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.
—— (2009), 'Hybrid Expressivism: Virtues and Vices', *Ethics*, 119: 257–309.
—— (2010), *Noncognitivism in Ethics*. New York: Routledge.
Unwin, N. (2001), 'Norms and Negation: A Problem for Gibbard's Logic', *Philosophical Quarterly*, 51: 60–75.
van Roojen, M. (1996), 'Expressivism and Irrationality', *Philosophical Review*, 105: 311–35.