Mark Schroeder
University of Southern California
January 3, 2014

the truth in hybrid semantics

forthcoming in Guy Fletcher and Michael Ridge, eds., *Having it Both Ways: Hybrid Theories and Modern Metaethics*

It is a familiar idea that the meanings of some words cannot be fully captured through a compositional characterization of truth-conditions. So far as its semantic contribution to truth-conditional compounds goes, 'but' is the same as 'and', but 'but' means something more than 'and', having a conventionally associated secondary content of some sort. It is intuitive to many that this same characterization is satisfied by racial and ethnic slurs, among a wide range of other words in natural language. Grice called this sort of meaning 'conventional implicature'.[1] With Grice as their forbear, and the example of racial and ethnic slurs as a model, a growing number of contemporary theorists[2] have proposed that distinctively moral vocabulary is special precisely because its meaningfulness exceeds core truth-conditional content. Such theorists offer *hybrid* accounts of moral language, on which a further, often noncognitivist, component is appended to a core, cognitivist account.

In my view, hybrid theories of this kind are primarily of interest not because they are a way of reviving the nondescriptivist or irrealist aspirations of traditional forms of noncognitivism, but instead because they offer a promising tool for naturalist or reductive realist theories. If moral facts or properties are simply natural properties, but moral words are loaded ways of speaking about those properties and moral thoughts are loaded ways of thinking about them, as the hybrid model suggests, then the hybrid naturalist can offer an account of what is distinctive and important about moral language and thought, despite the place of its subject matter in the natural world. At any rate, that is the background motivation for the kind of hybrid view in which I will be interested in this paper.[3]

---

[1] Grice [1989].

[2] See in particular, Barker [2000], Copp [2001], [2009], [this volume], Boisvert [2008], Strandberg [2011], and Hay [2011].

[3] Compare Copp [2001], Boisvert [2007] and Hay [2011], and contrast Ridge [2006], [2007]. Copp's motivations and commitments differ from those of Boisvert and Hay in an important way that I'll discuss in section 4. Tresan [2006], [2009] also sketches a view with many of the same features as that discussed here.

As I argued in an earlier paper[4], one of the most important issues facing hybrid metaethical theories concerns the interaction of moral vocabulary and attitude verbs. In order to attain any distinctive virtues not sharable by a similar purely cognitivist theory, I argued, the hybrid theorist must hold that the secondary content of moral terms, although it 'projects' through truth-conditional connectives, contributes to the primary, truth-conditional, content of attitude verbs – in particular, 'believes'. In contrast, I argued, the secondary content of racial slurs seems to project through attitude verbs, and does not seem to contribute to their truth-conditional content. So if moral vocabulary fits the hybrid theorist's needs, I argued, racial slurs aren't really such a close analogy after all.

This paper takes up where that earlier paper left off by presenting a simple, flexible model for secondary contents. This model shows clearly that there is nothing fundamentally strange about secondary contents affecting the core, 'truth-conditional', content of attitude verbs, and I'll argue that there are clear examples of this in natural language – though racial slurs are not among them. Then I'll take up the all-important question of how we should expect the word 'true' to work in a language which follows this flexible model. This will allow me to re-state the worry about whether moral vocabulary could really work as the hybrid theorist needs in a sharper way.

I   two patterns of projection

As already noted, there are a number of prominent examples of words whose conventional meanings appear to encode more than we think of as figuring into their core, 'truth-conditional', contents. This is often said, for example, not only about 'but' and 'even', but about racial and ethnic slurs, of which I'll take 'cheesehead' to be a mild exemplar for purposes of this paper.[5] These words share several important patterns. First, when speakers use simple sentences involving these words, they communicate more than one thing. A speaker who says, "Shaq is huge but agile" communicates both that Shaq is huge and agile, and also that she sees a contrast between being huge and being agile.[6] Similarly, a speaker who says, "the neighborhood is filling up with cheeseheads" communicates both that the neighborhood is filling up with people from Wisconsin, and also that she doesn't think very well of people from Wisconsin.

The reason why we don't take these second aspects of meaning to be 'truth-conditional', however, is that they seem to make a different contribution to so-called 'truth-conditional' connectives like negation

---

[4] 'Hybrid Expressivism: Virtues and Vices,' *Ethics*, reprinted in the 2009 *Philosophers' Annual*.
[5] See http://en.wikipedia.org/wiki/Cheesehead.
[6] This example comes from Bach [1999].

and disjunction. If a speaker says, "Shaq is not huge but agile," we read her as communicating not that either Shaq is not huge or he is not agile or she does not see a contrast between the two, but rather that he is not both huge and agile, and that she *does* see a contrast between the two. Similarly, if she says, "the neighborhood is not filling up with cheeseheads," we don't take her to have communicated that either the neighborhood is not filling up with people from Wisconsin or she does think well of people from Wisconsin, but rather that the neighborhood is not filling up with people from Wisconsin, and that she does not think very well of people from Wisconsin. So if negation operates on truth-conditions, then the extra semantic contribution of 'but' and 'cheesehead' does not contribute to truth-conditions.

Similar points apply to other constructions: someone who says "either Shaq is huge but agile or my eyes deceive me" is not committed to the claim that Shaq is both huge and agile, but she does communicate that she sees a contrast between being huge and being agile. Similarly, someone who says, "either the neighborhood is filling up with cheeseheads, or I still can't tell the difference between Wisconsin and Minnesota accents" is not committed to the claim that the neighborhood is filling up with people from Wisconsin, but she does communicate that she does not think very well of people from Wisconsin.

In general, the two intuitive parts of the meanings of 'but' and 'cheesehead' make distinct contributions to complex sentences formed by 'truth-conditional' contexts like negation and disjunction. One of these parts appears to be affected by the connectives in the familiar truth-conditional way, but the other appears to merely be 'passed up' to the more complex sentence. When a piece of information is inherited by a complex sentence directly from one of its parts, we may say, following the literature on presupposition, that it *projects*. What we have observed so far is that 'but' and 'cheesehead' appear to be conventionally associated with information that projects through constructions like negation and disjunction. It is natural to conjecture that this may be true of all complex sentences in which these terms figure – that the 'extra' semantic content associated with 'but' and 'cheesehead' will project through all ways of embedding them in more complex constructions. But that conjecture would be incorrect. Though it has some promise for the case of slurs, it is clearly incorrect for 'but'. The main source of the difficulty arises from attitude verbs.

To see this, first consider "Jill believes that the neighborhood is filling up with cheeseheads." Arguably, someone who thinks poorly of people from Wisconsin can use this sentence to report Jill's belief even if she recognizes that Jill is soft on people from Wisconsin, but in saying so, she communicates her own poor regard for people from Wisconsin. If that is right, then the extra content of 'cheesehead' can at least sometimes project through attitude verbs like 'believes' (though contrast 'said', which arguably behaves

differently). However, in contrast consider "Marv believes that Shaq is huge but agile." Someone who says this does *not* communicate that she sees any contrast between Shaq's being huge and his being agile, and she speaks falsely unless *Marv* sees such a contrast.[7] So the extra semantic content of 'but' does not project through 'believes'; on the contrary, it directly contributes to the truth-conditions of 'believes' reports.

This contrast is important, because as we'll see in what follows, it's important for the purposes of the hybrid theorist in whom I will be interested, that moral words like 'wrong' pattern with 'but', rather than with 'cheesehead', in this respect. This is because it is an important part of the prospect for the hybrid view to tell us something distinctive about the practical import of moral language and thought that sentences like "Karen believes that stealing is wrong" are true only if Karen has an appropriately related non-cognitive attitude.

Though slurs like 'cheesehead' contrast with 'but' in their projection behavior, it is important to distinguish both from presuppositions proper. Typically, when a sentence Q has a presupposition P, that presupposition will not be carried by the conditional sentence, if P, then Q. For example, 'the king of France has a mustache' presupposes that France has a king, but 'if France has a king, then the king of France has a mustache' does not. In other words, conditionals are *filters* for presuppositions, because whether the presupposition of the consequent passes to the whole sentence depends on what figures in the antecedent. By contrast, the additional content of 'Mark is a good driver but from Wisconsin' seems to be that most people from Wisconsin are not good drivers, but 'if most people from Wisconsin are not good drivers, then Mark is a good driver but from Wisconsin' does not seem to be free of the extra commitment carried by its consequent. And I cannot see how to construct a sentence of the form, 'if P, then Mark is a cheesehead' that does not implicate the speaker in disdain for people from Wisconsin. So it is best, I think, to think of each of these as different phenomena from that of presupposition, worth investigating in their own right.

2   big hypothesis semantics

In 'Hybrid Expressivism: Virtues and Vices' I argued that a promising hybrid theory needs the secondary content of moral words to interact with attitude verbs like 'believes' in the way that I've just argued that 'but' does. I called this the 'big hypothesis', and argued that the case of slurs should not give us any optimism that it is correct, since slurs do not, in fact, work in this way. But it's important not to

---

[7] Compare again Bach [1999].

misunderstand the force of this point. The fact that slurs do not work in this way does not show that moral terms cannot – it just shows that insofar as they do, slurs are only an imperfect analogy for moral terms.[8] Indeed, it is very easy to describe how a language could work in which there are both terms like 'but' whose secondary contents interact with attitude verbs and terms like 'cheesehead' whose secondary contents do not. All that we need to do, is to allow for two different kinds of secondary contents.

In this section I'll set out a simple, flexible framework for thinking about secondary contents. What makes this framework flexible is that it is specifically formulated in order to allow a distinction between secondary contents that project through attitude verbs (as for 'cheesehead') and those which instead contribute to their truth-conditions (as for 'but'). So in order to make this distinction, each sentence will be assigned to a *triple* of semantic values: a *core* content, a *primary set*, and a *secondary set*. The primary set and secondary set are each to be thought of as sets of 'extra' content possessed by the sentence, and for the purposes of this paper, I will assume that their contents are states of mind – mental properties. However, for simplicity I will take the core content to be a conventional proposition. It will be easy to adapt the framework offered here to change either of these two assumptions; indeed this point will be important, later.

So our semantics is going to work by assigning every sentence of the language to a triple – a *core* content, a *primary* set, and a *secondary* set. Let's say that $[S]_0^a$ denotes the core content associated with 'S' relative to assignment a, that $[S]_1^a$ denotes the primary set associated with 'S' relative to assignment a, and that $[S]_2^a$ denotes the secondary set associated with 'S' relative to assignment a. So we need our semantics to assign each well-formed formula 'S' to the triple, $\langle [S]_0^a, [S]_1^a, [S]_2^a \rangle$. Because only $[S]_0^a$ represents 'core' content, it is the only thing that we want our 'truth-conditional' connectives to substantively affect. With respect to such connectives, the primary and secondary sets will behave in the same way. But attitude verbs will treat them differently, allowing the primary set, but not the secondary set, to contribute to the core content of attitude ascriptions. Those are the ideas encapsulated in the following rules:

Connectives:
$[A \text{ and } B]_0^a = \text{TRUE iff } [A]_0^a = \text{TRUE and } [B]_0^a = \text{TRUE}$
$[A \text{ and } B]_1^a = [A]_1^a \cup [B]_1^a$
$[A \text{ and } B]_2^a = [A]_2^a \cup [B]_2^a$

---

[8] Ryan Hay [2011] argues persuasively that a more promising analogy for moral terms are what he calls 'general pejoratives' like 'jerk' and 'asshole'.

$[\sim A]_0^a = \text{TRUE iff } [A]_0^a \neq \text{TRUE}$
$[\sim A]_1^a = [A]_1^a$
$[\sim A]_2^a = [A]_2^a$

$[\forall x(A)]_0^a = \text{TRUE iff for every assignment b which differs from a at most in what it assigns to x,}$
$\qquad [A]_0^b = \text{TRUE}$
$[\forall x(A)]_1^a = [A]_1^a.$
$[\forall x(A)]_2^a = [A]_2^a.$

Attitude verbs:

$[\text{x believes that } A]_0^a = \text{TRUE iff o believes } [A]_0^a \text{ and for all } P \in [A]_1^a, \text{ o instantiates P, where a}$
$\qquad \text{assigns o to x.}$
$[\text{x believes that } A]_1^a = \varnothing, \text{ for all a.}$
$[\text{x believes that } A]_2^a = [A]_2^a.$

$[\text{x wonders whether } A]_0^a = \text{TRUE iff o wonders whether } [A]_0^a \text{ and for all } P \in [A]_1^a, \text{ o instantiates P,}$
$\qquad \text{where a assigns o to x.}$
$[\text{x wonders whether } A]_1^a = \varnothing, \text{ for all a.}$
$[\text{x wonders whether } A]_2^a = [A]_2^a.$

These rules distinguish between primary and secondary contents in precisely the way anticipated. So all it takes to see how these principles work in action is to introduce some vocabulary that contributes to the primary and secondary contents.

Predicates:

$[\text{x is from Wisconsin}]_0^a = \text{TRUE iff o is from Wisconsin, where a assigns x to o}$
$[\text{x is from Wisconsin}]_1^a = \varnothing, \text{ for all a}$
$[\text{x is from Wisconsin}]_2^a = \varnothing, \text{ for all a}$

$[\text{x can drive in traffic}]_0^a = \text{TRUE iff o can drive in traffic, where a assigns x to o}$
$[\text{x can drive in traffic}]_1^a = \varnothing, \text{ for all a}$
$[\text{x can drive in traffic}]_2^a = \varnothing, \text{ for all a}$

$[\text{x is a cheesehead}]_0^a = \text{TRUE iff o is from Wisconsin, where a assigns x to o.}$
$[\text{x is a cheesehead}]_1^a = \varnothing, \text{ for all a.}$
$[\text{x is a cheesehead}]_2^a = \{\text{DISDAIN(people from Wisconsin)}\}, \text{ for all a.}$

New Connective:

$[\text{A but B}]_0^a = \text{TRUE iff } [A]_0^a = \text{TRUE and } [B]_0^a = \text{TRUE}$
$[\text{A but B}]_1^a = [A]_1^a \cup [B]_1^a \cup \{\text{BF(A contrasts with B)}\}$
$[\text{A but B}]_2^a = [A]_2^a \cup [B]_2^a$

On this picture, the extra contents associated with 'but' and with 'cheesehead' behave differently with respect to attitude verbs. Someone who says 'Mark believes that Schroeder is a cheesehead' communicates her own disdain for people from Wisconsin, rather than telling us about Mark's disdain, but someone who says 'Mark believes that Schroeder is from Wisconsin but Schroeder can drive in traffic' communicates that Mark, rather than the speaker, sees a contrast between being from Wisconsin and being able to drive in traffic. And finally, someone who says 'Caroline believes that Schroeder is a cheesehead but can drive in traffic' communicates that he, rather than Caroline, has disdain for people from Wisconsin, but that Caroline, rather than he, sees a contrast between being from Wisconsin and being able to drive in traffic.

We can think of the framework described here as a minimal model for marking the distinction between terms which behave in these two different ways. The model is highly flexible, both in that it leaves open that attitude verbs other than 'believes' may treat the primary and secondary contents differently, and that it allows, in principle, for words that affect both the primary and secondary contents. It's also easy to use it to implement other interesting views. For example, Ryan Hay's view about the semantics of 'jerk' can be implemented (roughly) as follows:[9]

Jerk:

$[x \text{ is a jerk}]_0^a = \text{TRUE}$ iff o is exhibits sufficient disregard for others, where a assigns x to o.
$[x \text{ is a jerk}]_1^a = \{\text{RESENTMENT}(\text{people who exhibit sufficient disregard for others})\}$, for all a.
$[x \text{ is a jerk}]_2^a = \varnothing$, for all a.

It is easy to use our compositional principles to derive the conclusion that someone who says 'Caroline believes that not everyone who believes no cheeseheads can drive in traffic is a jerk' communicates that Caroline resents people with sufficient disregard for others without expressing this attitude herself, but does express her own disdain for people from Wisconsin.

3   logic and inference

The foregoing framework makes it easy to observe two important lessons about hybrid metaethical theories: first, that for purposes of understanding logic and inference, it does not matter whether the 'extra' content of moral vocabulary is primary or secondary content; but second, that for purposes of attaining the distinctive advantages of hybrid theories over ordinary cognitivism, it is important that moral terms

---

[9] Officially, Hay [2011] rejects the idea that there is any straightforward paraphrase of the core content of 'jerk', and the formulation here is more concrete about the associated attitude than Hay is, but this is essentially the shape of his picture.

contribute to primary content. I'll take these two lessons one by one, focusing on logic and inference in this section, and on the distinctive advantages over cognitivism in the next.

The problem of accounting for logic and inference can be broken down into two parts: those concerned with *inconsistency*, and those concerned with *inference*.[10] Inconsistency is the easier of these two issues, because it can be fully accounted for in terms of core contents, without the need to worry about primary or secondary contents. Our compositional rules guarantee that any two sentences are logically inconsistent just in case no interpretation could make their core contents both true. But inference is potentially harder to explain. In general, logically valid arguments should commit a thinker who endorses their premises to endorsing their conclusion. But since some sentences are associated with primary and secondary attitudes, endorsing the conclusion of a valid argument often involves having such further states of mind, in addition to the belief in the core content. For example, endorsing 'P and Q' should not commit a thinker to endorsing 'P but Q'. It follows that understanding how endorsing the premises of a valid argument can endorse a thinker to endorsing their conclusion requires paying attention to primary and secondary sets.

Fortunately, both primary and secondary sets have exactly the feature that is required in order to solve this problem. I'll first illustrate with two examples that we've already introduced, and then generalize. First, consider a simple *modus ponens* argument involving 'cheesehead':

P1   Mark cannot drive in traffic
P2   If Mark cannot drive in traffic, then Mark is a cheesehead
C1   Mark is a cheesehead

Endorsing the premises of this argument should transparently commit someone to endorsing its conclusion (though of course, this commitment can be given up by giving up one of the premises). The conclusion has the core content that Mark is from Wisconsin, and has the secondary set consisting of disdain for people from Wisconsin. Endorsing the conclusion therefore requires both believing this primary content and also having disdain for people from Wisconsin. And so a thinker is committed to endorsing the conclusion only if she is committed to both this belief and this disdain. There is no puzzle where the commitment to this belief comes from, because endorsing the premises of the argument requires the thinker to believe their core contents, and it's a familiar observation that having beliefs commits you to believing their consequences. But similarly, there is no puzzle where the commitment to the disdain comes from,

---

[10] For further discussion, see Schroeder [2009], [2010, chapter 6].

because it is also part of the secondary set for premise P2. That means that if the thinker endorses the premises of the argument, then she *already* has disdain for people from Wisconsin.

The exact same explanation goes, for a *modus ponens* argument involving a non-vacuous primary set (our examples could be more interesting if our language was richer):

> **P3** Mark is from Wisconsin
> **P4** If Mark is from Wisconsin, then Mark is a jerk
> **C2** Mark is a jerk

C2 is associated with a primary set including resentment of people who have sufficient disregard for others, and so to endorse it, a thinker must be in that state of resentment, in addition to believing its core content. But this state of resentment is also part of the primary content of the second premise. And so a thinker cannot endorse both of the premises without being in this state. So any thinker who endorses both premises is already in all of the states that are in the primary set for the conclusion. But believing the core contents of the premises commits a thinker to believing the core content of the conclusion. So a thinker who endorses the premises of this argument is committed full-stop to endorsing its conclusion (though again, she can give this commitment up by giving up one of the premises).[11]

In general, productive, relevant validities of predicate logic are generally going to have this property – that states of mind which appear in either the primary or the secondary set for the conclusion also appear in the primary or secondary set for one of the premises. This principle only fails when the conclusion introduces new terminology that doesn't appear in one of the premises. So, for example, endorsing the premises of the following argument does not commit you to endorsing its conclusion:

> **P5** Mark cannot drive in traffic
> **C3** Either Mark cannot drive in traffic, or he is a cheesehead

Though endorsing the premise of this argument commits you to believing the core content of its conclusion, it does not commit you to endorsing the conclusion full-stop, because to endorse the conclusion full-stop, you must have disdain for people from Wisconsin, but nothing about accepting the premise commits you to this disdain. But this is the right prediction. If 'cheesehead' is a slur, thinkers should not be committed to endorsing sentences involving it, even by vacuous inferences like disjunction introduction. So I conclude that the framework described here accommodates just the right amount of

---

[11] Compare Boisvert [2007], Schroeder [2009].

inference. The same will go for moral terms, if they follow either the model of 'jerk' and affect the primary set, or the model of 'cheesehead' and affect the secondary set – or even if they affect both. So the distinction between primary and secondary sets is not relevant for accounting for logic and inference – at least, not the inferences that we want, from ordinary predicate logic.[12]

4   advantages over cognitivism

In contrast, the difference between primary and secondary contents is crucial for making good on what I take to be the most significant prospects for a hybrid metaethical theory to have important advantages over ordinary cognitivist views. This is because though it follows from both the view that moral language contributes to the primary set and the view that moral language contributes to the secondary set that moral *language* is special, only on the former view does it follow that moral *thought* is special. And a hybrid theory that makes sense of the specialness of moral language without making sense of the specialness of moral thought will be very limited in what advantages it can claim.

Consider, for example, the case of David Copp's realist expressivism.[13] According to Copp, moral language is special, because moral sentences *conventionally simplicate* that the speaker has a corresponding attitude. But Copp denies that having this attitude is required for having moral beliefs. So according to Copp, you can have moral beliefs without having any attitude that would motivate you to act, as a result. It is not surprising that Copp should defend such a view, because he has long advocated judgment externalism about moral judgments. But this leads to the question as to what Copp gains by hybridizing his view that he could not have within an ordinary cognitivist view. According to Copp, what he gains is an explanation of why there is an *illusion* of judgment internalism, on the grounds that when speakers overtly assert moral sentences, we expect them to be motivated, because those sentences conventionally simplicate that the speaker has the attitude required in order for her to be motivated.[14]

But this does not strike me as a particularly compelling explanation of why moral judgment internalism would seem to be true. Judgment internalism, after all, does not say only that speakers who *assert* moral claims can be expected to be motivated, but that people who genuinely *believe* moral claims will be motivated. It may help to compare an example which may plausibly be treated as affecting secondary content, the formal/informal distinction for the second-person pronoun. Complicating the simple

---

[12] I'll turn to a different pattern of inference, involving the word 'true', in section 5.

[13] Copp [2001], [2009], [this volume].

[14] "Second, it can explain the intuitions that lead people to accept moral judgment internalism (or judgment motivational internalism) in a way that is compatible with externalism" [this volume, MS p7].

language that we've been describing enough to formally incorporate the formality distinction for second-personal pronouns would take us further astray than the resulting payoff would justify, but the idea is that second-personal pronouns like 'vous'/'tu', 'Sie'/'du', and the now-dead 'you/thou' each make a contribution to a sentence's secondary set, intimating an expectation either of familiarity or of distance on the part of the speaker toward her addressee. It makes sense to incorporate this into the secondary set, rather than the primary set, because when a speaker says, 'Julia glaubt, daβ du Deutsch sprichst', he communicates, not that Julia expects familiarity with his addressee, but that he does.[15]

The formal/informal distinction makes for a useful test of Copp's hypothesis of how there could be an illusion of judgment internalism, because we *do* expect someone who asserts 'Du sprichst Deutch' to expect familiarity with her audience, but this wouldn't lead us in any way to suspect, as we might put it, that jemand, der glaubt, daβ du Deutch sprichst, would expect to be familiar with you. Nor would we describe someone who did suspect this as succumbing to a natural but misleading intuition. Since there is no illusion in the formal/informal case, that leads me to doubt how there could be a similar illusion in the moral case.

I conclude that hybridization offers much deeper prospects to explain something fundamentally distinctive about morality only if it promises to tell us something not only about moral language, but about moral thought. And it is only on the view that moral language affects *primary* content that this is so. The reason for this is simple. We can only succeed in saying that there is something special about moral belief, using the sentence, 'there is something special about moral belief', if the meaning of the word 'belief' allows us to say this. So whatever we say about moral belief needs to be constrained by the semantics for 'belief' or 'believes'.[16]

The distinction between primary and secondary sets makes all the difference, for whether the distinctive extra feature of moral language affects the content of claims of the form, 'x believes that stealing is wrong'. So it is what allows us to coherently use the word 'believes' to say that something more is required in order to believe that stealing is wrong, than to believe its core content. Here is a simple example of such a view:

---

[15] For a wide collection of examples of words and constructions that clearly work along the model of secondary sets, including discussion of the formal/informal distinction, see Potts [2005].

[16] Observe that when Copp [this volume] tries to extend his view in order to argue that there is something distinctive of moral thought, he is forced to use quotation, and talk about what is required in order to '*think* that Brenda is a "pom" ' [this volume, 8 of MS].

Wrong, first pass:

$[$x is wrong$]_0{}^a = $ TRUE iff o fails to maximize happiness, where a assigns x to o.

$[$x is wrong$]_1{}^a = \{$DESIRE(not to do things that fail to maximize happiness)$\}$, for all a.

$[$x is wrong$]_2{}^a = \varnothing$, for all a.

On this simple picture, 'Jim believes that stealing fails to maximize happiness' can be true even if 'Jim believes that stealing is wrong' is not. So 'Jim believes that stealing fails to maximize happiness but Jim does not believe that stealing is wrong' can be true. So, semantically descending, we can agree that Jim believes that stealing fails to maximize happiness but does not believe that stealing is wrong. Hence, on this view, we can make sense of the idea that there is more to moral belief than belief in some ordinary, natural fact.[17]

On the particular view that I've just described, the "more" that there is to moral belief is exactly what would be needed for it to turn out that judgment internalism is true, even if ordinary beliefs never motivate by themselves, but only, as the Humean Theory of Motivation claims, in connection with a desire. This is because it tells us that 'Jim believes that stealing is wrong' is not true unless Jim believes that stealing fails to maximize happiness and desires not to do things that fail to maximize happiness. Hence, semantically descending again, it tells us that Jim does not believe that stealing is wrong unless Jim has both this belief and this desire. But the combination of this belief with this desire is exactly what we would expect to motivate Jim not to steal. Hence, placing the extra content of 'wrong' into the primary set is exactly what we would need, in order to validate judgment internalism.

Yet a hybrid theory could also preserve the prediction that there is more to moral belief than any ordinary descriptive belief, *without* endorsing any strong form of judgment internalism.[18] To do so, we just need to change what we put in the primary set for 'wrong' sentences. Here is one such alternative:

Wrong, second pass:

$[$x is wrong$]_0{}^a = $ TRUE iff o fails to maximize happiness, where a assigns x to o.

$[$x is wrong$]_1{}^a = \{$DISAPPROVAL(things that fail to maximize happiness)$\}$, for all a.

$[$x is wrong$]_2{}^a = \varnothing$, for all a.

If it is possible to believe that stealing is F, and to disapprove of F things, without being motivated not to steal, then this version of the hybrid theory will not validate judgment internalism about 'wrong' beliefs.

---

[17] And by this we mean not only that there is more to the belief that something is "wrong" than belief in some natural fact, as Copp [this volume] can allow, but that there is more to the belief that it is wrong than belief in some natural fact. This further claim is not available to Copp, since he rejects the Big Hypothesis.

[18] Compare Hay [this volume].

But it will maintain the feature – preserved by all views on which 'wrong' contributes to the primary set – that there is more to moral beliefs than simply beliefs in natural, descriptive facts. This is, I think, a very intelligible advantage to be interested in over non-hybrid naturalist or reductive views.

As we've just seen, hybrid theory that holds that 'wrong' contributes to the primary set can, but does not need to, validate the truth of judgment internalism about 'wrong'. This version of judgment internalism is what Jon Tresan [2006] calls *de dicto* internalism, according to which necessarily, someone who believes that stealing is wrong will be motivated not to steal. Tresan contrasts de dicto internalism with *de re* internalism, according to which the belief that stealing is wrong is such that is necessarily motivates someone not to steal. Tresan advocates a view on which de dicto internalism is true, but de re internalism is false. On Tresan's view, the belief that stealing is wrong is just an ordinary descriptive belief that could exist in the absence of any motivation, but it does not count as a moral belief, or as the belief that stealing is wrong, unless that motivation is present.

The hybrid theory in which I'm interested in this paper is consistent with, but does not entail, Tresan's de dicto internalist cognitivism. Whether we develop the hybrid view in a way that leads to Tresan's view, or in a way that also endorses *de re* internalism, depends on what we take phrases like 'the belief that stealing is wrong' to refer to. Tresan's suggestion that these phrases refer to the ordinary descriptive belief that stealing fails to maximize happiness, but do so by a description that is not satisfied unless a further motivation is present, is consistent with everything that I've said so far. On this interpretation, the hybrid theory offers an explanation of some of the puzzling features of Tresan's view, which he himself acknowledges.[19]

---

[19] In fact, however, I think it is more natural for the hybrid theorist to identify the belief that stealing is wrong with the conjunctive state that figures in the primary content of 'x believes that stealing is wrong'. And that is the way that I interpret the hybrid view in Schroeder [2013]. The resulting view shares many of the advantages carefully documented by Tresan [2006], including maintaining that representational and motivational states are modally separable, but parts from his view in denying that 'belief', in natural language, picks out the modally separable state that is exhausted by its representational character. This allows the view to maintain, contra Tresan, that the belief that stealing is wrong cannot exist without being the belief that stealing is wrong.

However, Tresan's view has at least a couple of advantages. Because he isn't committed to the background condition on a belief counting as a moral belief being a purely psychological condition, he can avail himself of a wider range of such background conditions than can plausibly be identified as part of the belief itself. For example, following suggestions by Foot [1978b] and Dreier [1990] in his [2006] he explores the possibility of social conditions. He also suggests that de dicto internalist cognitivism can mix and match any cognitivist and expressivist views.

5   truth in hybrid semantics

So far in this paper, I've given a simple model for how a 'big hypothesis' semantic framework might work – one that allows for a distinction between two different ways in which the conventional meaning of a word can exceed the contribution that it makes to the so-called 'truth-conditional' contexts of negation, conjunction, and disjunction. This distinction, which I've called the distinction between the 'primary' and 'secondary' sets associated with a sentence, allows us to distinguish between words whose extra conventionally associated character contributes to what it takes to have the associated beliefs, and those which do not. Arguably, 'but' and 'cheesehead' fall on opposite sides of this distinction in ordinary English, and I've argued that the advantages of adopting a hybrid theory that holds that 'wrong' patterns with 'cheesehead' are illusory. The most that we can get from such a hybrid theory is that there is something distinctive of moral *assertions*, but not that there is anything – or even an illusion of there being anything – distinctive about the thought that something is wrong. In order to hold that there is something distinctive of the thought that something is wrong, the hybrid theorist must hold that 'wrong' patterns with 'but', so that it can affect the semantics of thought-words like 'thought' and 'believes'. To say this is just to semantically ascend from the claim that there is something distinctive of beliefs and thoughts about what is wrong.

I now turn to the question of how the word 'true' fits into our semantic picture. In particular, how should 'true' interact with the primary and secondary sets associated with sentences? Pictures like the one described in the last few sections are often described as ones on which meaning exceeds truth-conditional content. On this view, only core content is truth-conditional, which suggests that 'true' and 'false' should only report on core content. This suggests, at a first pass, the following simple picture for 'true':

$$[\text{it is true that P}]_0^a = \text{TRUE iff } [\text{P}]_0^a = \text{TRUE}$$
$$[\text{it is true that P}]_1^a = \varnothing$$
$$[\text{it is true that P}]_2^a = \varnothing$$

This picture yields the strange conclusion that you can assert 'it is true that Mark is a cheesehead' without implicating yourself in disdain for people from Wisconsin, or assert 'it is true that Mark is from Wisconsin but can drive in traffic' without implicating yourself in any contrast between being from Wisconsin and being able to drive in traffic. But both of these seem wrong.

But perhaps the idea that only core content is 'truth-conditional' should not be identified with the idea that 'true' does nothing other than to report on core content. Perhaps it should instead be identified

with the idea that the only thing 'true' reports on, is core content. In other words, perhaps in addition to reporting on core content, sentences involving 'true' may also, at least in some cases, have non-empty primary or secondary sets, because 'true' does not block the projection of primary or secondary sets. Here is what such a view would look like:

$$[\text{it is true that P}]_0^a = \text{TRUE iff } [\text{P}]_0^a = \text{TRUE}$$
$$[\text{it is true that P}]_1^a = [\text{P}]_1^a$$
$$[\text{it is true that P}]_2^a = [\text{P}]_2^a$$

This view makes perfect sense of why you would only assert 'it is true that Mark is a cheesehead' only if you would also be willing to assert 'Mark is a cheesehead', and similarly for 'it is true that Mark is from Wisconsin but can drive in traffic'.[20] Moreover, in company with our clause for 'believes', it makes sense of why minimally reflective agents will believe that it is true that Mark is from Wisconsin but can drive in traffic just in case they believe that Mark is from Wisconsin but can drive in traffic. These seem like plausible predictions.

These predictions are plausible enough for sentences of the form 'it is true that P'. But most interesting sentences containing the word 'true' do not have this form. After all, if 'it is true that P' and 'P' are equivalent, then we can get by just as well using only the latter, as using the former. So interesting sentences involving 'true' say things like 'what Caroline believes is true' and 'everything that cheesehead said is true'. We can use such sentences and accept them even if we do not know what Caroline believes or everything that the person being referred to as a 'cheesehead' said.

Paying attention to these sentences helps us to zero in on which words in 'it is true that Mark is a cheesehead' are responsible for the contents of its secondary set. This is because 'what Caroline believes is true' does not implicate the speaker in disdain for people from Wisconsin, even if it happens to be the case that Caroline would express her belief using the sentence, 'Mark is a cheesehead'. For the speaker of 'what Caroline believes is true' might not even know what Caroline believes about Mark. In contrast, 'everything that cheesehead said is true' does implicate its speaker in disdain for people from Wisconsin — even if the person being referred to as a 'cheesehead' does not have any beliefs they would express by using the word 'cheesehead'. So in both the case of 'it is true that Mark is a cheesehead' and 'everything that cheesehead

---

[20] Contra Bach [1999], this view preserves the thesis that conventional implicatures like that of 'but' do not contribute to truth-conditions, while honoring the idea that when we say that it is true that Shaq is huge but agile, we are also committing to the claim that it is true that there is a contrast between being huge and being agile. But on this view, this is because the extra conventionally associated content projects upwards through 'true', rather than because it contributes to the core content of 'true' sentences.

said is true', the secondary set contains disdain for people from Wisconsin, but in neither case is this contributed by the word 'true', any more than it is contributed by the word 'or' in 'either Mark is a cheesehead or he is not from Wisconsin'.[21] Instead, it is contributed by the word 'cheesehead', and merely projected upward through the word 'true'. So what this tells us, is that secondary sets project upwards through 'true', just as they project through 'truth-conditional' connectives and attitude verbs like 'believes'. This should not be a surprise.[22]

The fact that 'what Caroline believes is true' does not implicate its speaker in disdain for people from Wisconsin has important consequences for the following argument:

> **P6**   What Caroline believes is true.
> **P7**   What Caroline believes is that Mark is a cheesehead.
> **C4**   Mark is a cheesehead.

This argument is not only transparently valid, but also clearly has the property that accepting the premises commits you, on pain of giving one of them up, to accepting the conclusion. But accepting the conclusion requires having disdain for people from Wisconsin. What ensures that someone who accepts the premises will have or be committed to having such disdain? I've just been arguing that 'what Caroline believes is true' does not implicate the speaker in such disdain. So it doesn't come from accepting the first premise. But fortunately, our semantic clause for 'believes' tells us that 'Caroline believes that Mark is a cheesehead' *does* implicate the speaker in such disdain. So sentences with secondary sets validate the inference schema, 'what Caroline believes is true'; 'what Caroline believes is that P': 'P', and they do so precisely because secondary sets project through attitude verbs.

If this is right, then it should make us worried about the idea that moral words are associated with *primary* sets, rather than *secondary* sets. For the following argument, like the previous one, is not only transparently valid, but clearly has the property that accepting the premises commits you, on pain of giving one of them up, to accepting the conclusion:

> **P8**   What Caroline believes is true.
> **P9**   What Caroline believes is that stealing is wrong.
> **C5**   Stealing is wrong.

---

[21] Contrast Ridge [2009], on whose view all uses of 'true' express the very same attitude that is expressed by normative words like 'wrong' and 'reason'.

[22] It would be easy to amend our semantics to show how this works, and to allow for the construction of sentences like P7 and P9, by replacing our operator treatment of 'believes' with a clause that treats 'believes' as expressing a relation between agents and the semantic values for sentences. I won't carry out this exercise, here.

But as we've seen, you can accept a sentence like P8 without having any particular attitude, and in particular, without disapproving of things that fail to maximize happiness. And on the assumption that 'stealing is wrong' has a non-empty *primary* set, rather than a non-empty *secondary* set, it follows that you can also accept P9 without having any particular attitude, and in particular, without disapproving of things that fail to maximize happiness. But in order to accept C5, you need to have some further attitude. On our second-pass gloss at a primary-set hybrid view of 'wrong', it is that you must disapprove of thing that fail to maximize happiness. And so you are *committed* to accepting C5 by accepting P8 and P9 only if accepting them commits you to this state of disapproval. But now we have no explanation – or at least, none on a par for how things worked for the productive validities of classical logic or for the argument from P6 and P7 to C4 – of why this would be so. This looks like a *problem* for the view that 'wrong' contributes to the primary set, rather than to the secondary set.

You might be suspicious that something has gone wrong in the setup of this problem. For all along in this paper, I've worked with a working example of a word that I've been suggesting *does* contribute to the primary, rather than the secondary, set: 'but'. So shouldn't the foregoing considerations lead us to predict that the following argument should not have the property that accepting the premises commits you, on pain of giving one of them up, to accepting its conclusion?

P10 What Caroline believes is true.
P11 What Caroline believes is that Mark is from Wisconsin but can drive in traffic.
C6 Mark is from Wisconsin but can drive in traffic.

If my foregoing remarks are on the right track, then you can accept P10 without believing that there is a contrast between being from Wisconsin and being able to drive in traffic. And if 'but' really does contribute to the primary but not the secondary set, as I've argued, then you can accept P11 without believing that there is a contrast between being from Wisconsin and being able to drive in traffic. So shouldn't there be a puzzle about how accepting P10 and P11 commits you to accepting C7, which *does* require you to believe that there is a contrast between being from Wisconsin and being able to drive in traffic? But this argument *does* seem to commit someone who accepts its premises to accepting its conclusion.

There is an important difference, however, between the case of 'but' and the case of 'wrong', at least as the hybrid theorist conceives of it. And that is because the primary set for 'but' includes a *belief*. So when you accept P11, you believe not only that Caroline believes that Mark is from Wisconsin and can

drive in traffic, but also that Caroline believes that there is a contrast between being from Wisconsin and being able to drive in traffic. So if you also accept P10, and think that what Caroline believes is true, then you are committed to thinking that it is true that there is a contrast between being from Wisconsin and being able to drive in traffic, and hence that there is such a contrast. So you really *are* committed to the primary set for C7, in virtue of accepting both P10 and P11 – and the reason you are so committed is that the primary set of P11 is just another belief.

In contrast, on the hybrid view, the primary set for 'wrong' is something other than another run-of-the-mill belief. For if it were just another run-of-the-mill descriptive belief, then moral belief would not turn out, after all, to be something over and above ordinary, run-of-the-mill descriptive belief. On this view, both the argument from P6 and P7 to C4 and the argument from P10 and P11 to C6 commit someone who accepts their premises to accepting their conclusion, but they do so for different reasons – and neither of these reasons extends to the argument from P8 and P9 to C5, on the hybrid view under consideration. So the case of 'but' should not make us optimistic; on the contrary, it strongly suggests that the only things that can figure in primary sets are further beliefs.

6   seizing both horns?

In part 4 I argued that in order to be able to say that moral belief is something over and above ordinary descriptive belief, the hybrid theorist must say that moral words contribute to primary sets. But in part 5 I've argued that in order to make sense of the way in which simple arguments involving 'true' preserve commitment, the hybrid theorist must not say that moral words contribute *only* to primary sets. But we can now put these two conclusions together in order to argue that an adequate hybrid theory should say that moral words must contribute to *both* primary and secondary sets.

The argument is simple. By the considerations from part 4, we can say that there is something distinctive about moral thought only if the primary set for moral sentences such as 'stealing is wrong' is non-empty. This is what allows us, speaking the language, to coherently say such things as that it is possible to have any set of non-moral beliefs without believing that stealing is wrong. So without loss of generality, let us call the contents of the primary set for 'stealing is wrong' A. It follows that in order to be committed to accepting the conclusion of the following argument, you must be committed to having attitude A:

P8   What Caroline believes is true.
P9   What Caroline believes is that stealing is wrong.
C5   Stealing is wrong.

But P8 does not have A in either its primary or its secondary set. So absent some different, non-hybrid-style, explanation of how accepting the premises commits you to having A, the only way for accepting the premises to guarantee that you have A, is if A is in either the primary set or the secondary set for P9. But by our semantic clause for 'believes', this is true only if A is in the secondary set for 'stealing is wrong'. Consequently, 'stealing is wrong' must have a non-empty primary set, and everything in its primary set must also be in its secondary set.

It could certainly be that moral words work like this. But I find the argument that leads to this conclusion troubling. For it strikes me that it is no coincidence that arguments of the pattern of P6 and P7 to C4, P8 and P9 to C5, and P10 and P11 to C6 all have the property that accepting the premises commits you to accepting their conclusions. I suspect that we will find that *all* arguments with this form will have this 'inference-licensing' property. And that means that one of two things must be true. Either primary sets *never* contain attitudes other than ordinary descriptive beliefs, or else when they do, their non-belief members must *always* also be in the secondary set. But I find the latter of these two possibilities very hard to credit. We know from the example of 'but' that there are some words that contribute to the primary set but not the secondary set. So what possible mechanism could *prevent* words from doing this, when what they contribute to the primary set is something other than a belief?

In contrast, the hypothesis that primary sets only include beliefs is much more plausible. On this view, there is no need to explain what ensures that non-belief members of primary sets also need to be members of secondary sets, even though in general, a state of mind can be in a primary set without being in a secondary set. In fact, on this view it is plausible to treat the contents of primary sets as *propositions*, rather than as mental states, and to alter the clause for 'believes' so that it requires belief in each of the members of the primary set, as well as in the core content.

7   wrapping up

In this paper I've shown that far from being difficult to imagine how moral words could obey the principle that I've called the Big Hypothesis, it is easy to describe systems on which this is so, and indeed that there are plausible models for this in natural language. However, I've tried to put pressure on whether this is the right picture for moral language by looking in greater detail at how to think about how the word 'true'

works. This pressure is not perfectly sharp. It can be resisted by moving to hybrid views like Copp's realist expressivism, on which the advantages over corresponding cognitivist theories are even subtler. It can be resisted by insisting that moral words really do contribute to both primary and secondary sets. And for all that I have said here, it may be resistible by adopting yet a different view of how 'true' interacts with primary and secondary sets – one which gives up the idea that 'core' content counts as characterizing 'truth-conditions', on any natural sense. Still, despite these caveats, I do think that it should make us cautious about drawing quick conclusions about how much help the hybrid maneuver may be to realist naturalism in metaethics.[23]

references

Bach, Kent [1999]. 'The Myth of Conventional Implicature.' *Linguistics and Philosophy* 22(4): 327-366.

Barker, Stephen [2000]. 'Is Value Content a Component of Conventional Implicature?' *Analysis* 60(3): 268-279.

Boisvert, Daniel [2008]. 'Expressive-Assertivism.' *Pacific Philosophical Quarterly* 89(2): 169-203.

Copp, David [1995]. *Morality, Normativity, and Society.* New York: Oxford University Press.

_____ [2001]. 'Realist Expressivism: A Neglected Option for Moral Realism.' *Social Philosophy and Policy* 18: 1-43.

_____ [2009]. 'Realist-Expressivism and Conventional Implicature.' *Oxford Studies in Metaethics* 4: 167-202.

_____ [this volume]. 'Can a Hybrid Theory Have it Both Ways? Moral Thought, Open Questions, and Motivation.' Guy Fletcher and Michael Ridge, eds., *Having it Both Ways: Hybrid Theories and Modern Metaethics.* Oxford: Oxford University Press, pp-pp.

Dreier, James [1990]. 'Internalism and Speaker Relativism.' *Ethics* 100(1): 6-26.

Foot, Philippa [1978a]. *Virtues and Vices.* Los Angeles: University of California Press.

_____ [1978b]. 'Approval and Disapproval.' In Foot [1978a].

Grice, H.P. [1989]. *Studies in the Way of Words.* Cambridge, MA: Harvard University Press.

Hay, Ryan [2011]. 'Hybrid Expressivism and the Analogy Between Pejoratives and Moral Language.' *European Journal of Philosophy* 21(3): 450-474.

_____ [this volume]. 'Hybridity, Entangling, and Type-Generality: Attitudinal Requirements for Thinking Morally.' In Guy Fletcher and Michael Ridge, eds., *Having it Both Ways: Hybrid Theories and Modern Metaethics.*

Potts, Christopher [2005]. *The Logic of Conventional Implicatures.* New York: Oxford University Press.

Ridge, Michael [2006]. 'Ecumenical Expressivism: Finessing Frege.' *Ethics* 116(2): 302-336.

_____ [2007]. 'Ecumenical Expressivism: The Best of Both Worlds?' *Oxford Studies in Metaethics* 2: 51-76.

_____ [2009]. 'The Truth in Ecumenical Expressivism.' In David Sobel and Stephen Wall, eds., *Reasons for Action.* Cambridge: Cambridge University Press, 219-242.

Schroeder, Mark [2009]. 'Hybrid Expressivism: Virtues and Vices.' *Ethics* 119(2): 257-309.

_____ [2010]. *Noncognitivism in Ethics.* New York: Routledge/Taylor and Francis.

_____ [2013]. 'Tempered Expressivism.' *Oxford Studies in Metaethics* 8: 283-314.

Strandberg, Caj [2009]. 'The Pragmatics of Moral Motivation.' *Journal of Ethics* 15(4): 341-369.

Tresan, Jon [2006]. 'De Dicto Internalist Cognitivism.' *Nous* 40(1): 143-165.

_____ [2009]. 'Metaethical Internalism: Another Neglected Distinction.' *Journal of Ethics* 13(1): 51-72.