Semantics, moral
forthcoming in The International Encyclopedia of Ethics

Semantics is the investigation of meaning, and semantic theories, including semantic theories about moral language, come in two very different kinds. Descriptive semantic theories are theories about what words mean. So descriptive moral semantic theories are theories about what moral words mean: words like ‘good’, ‘better’, ‘right’, ‘must’, ‘ought’, ‘reason’, and ‘rational’. In contrast, foundational semantic theories are theories about why words mean what they do, or more specifically, about what makes it the case that words mean what they do. So foundational moral semantic theories are theories about what makes it the case that moral words mean what they do: words like ‘good’, ‘better’, and so on. Since both kinds of theory are sometimes referred to as varieties of ‘moral semantics’, this article will cover both.

Just as semantic theories themselves come in two kinds, descriptive semantics raises two very different kinds of issues. Some issues in descriptive semantics turn on the question of what kind of theory is required, in order to be able to account for the meanings of words – including of moral words, in particular. Such questions are questions about the nature of semantic theorizing. So, for example, according to truth-conditional theories, an adequate descriptive semantic theory needs to provide a recursive and compositional characterization of values for sentences which determine the truth conditions of each sentence. Whereas according to expressivist theories, an adequate descriptive semantic theory does not need to determine truth conditions at all, but does need to recursively and compositionally associate each sentence ‘P’ with a mental state – intuitively, with the state which constitutes what it is to think that P. Truth-conditional and expressivist theories involve very different conceptions of the nature of linguistic meaning, but their contrast is one of the most important issues in moral semantics, and they are not even the only live competitors.

In contrast, other issues in descriptive semantics, rather than turning on what kind of theory is required, raise questions about specific properties of the meanings of particular words. Some of these questions are theory-internal – for example, the question of which attitude is involved with thinking that
murder is wrong is internal to expressivism, and the question of what the conditions are under which
‘murder is wrong’ is true is internal to truth-conditional semantics. But other questions about the meanings
of particular words are theory-neutral, in the sense that they come up, no matter what kind of semantic theory
is correct. An important class of such questions includes questions about the number and categories of the
argument places associated with some word. For example, while some philosophers believe that ‘good’ is
semantically a one-place predicate, others believe that it is semantically a two-place relation, or even that it
is a three- or four-place relation. And of those who believe that it is a one-place relation, there are
differences over whether its argument place is filled by propositions, or events, or by items belonging to
some other category. Similarly, those who adopt two-or-more-place theories about ‘good’ can disagree
about the categories to which each of its argument places belong. And in addition, there are disagreements
about whether there is only one sense of ‘good’ or whether there is more than one, and if so, which ones
there are, how many argument places they have, and what categories those argument places belong to. All
of these are theory-neutral questions about the meaning of ‘good’, because they are questions that any
descriptive semantic theory about ‘good’ will need to answer, whether it is truth-conditional, expressivist, or
of some other variety.

In this article, we will introduce foundational theories of moral semantics in section 1, theories
about the nature of moral semantics in section 2, and theory-neutral questions about the semantics of

1 foundational semantic theories
As noted above, foundational semantic theories seek to answer the question of what makes it the case that
words mean what they do, or to put it in other words, in virtue of what words mean what they do. So a
foundational moral semantic theory seeks to answer the question of what makes it the case that moral words
mean what they do. Of course, the very idea of a foundational moral semantic theory raises the question of
just why we need a special theory at all of what makes it the case that moral words mean what they do.
After all, if moral words mean what they do in virtue of the same things that make any words mean what
they do, then there is no distinctive question for foundational moral semantic theories to ask, that is not
already asked by foundational semantic theories in general.

In keeping with this observation, foundational moral semantic theories split over a central question
from descriptive moral semantics: namely, whether moral words have the same kind of meaning as most
ordinary non-moral words. Some theorists, who believe that moral words have a different kind of meaning
from that of ordinary non-moral words, believe that a different sort of foundational semantic theory is required for moral words than for ordinary descriptive words. Whereas theorists who believe that moral words have the same kind of meaning as ordinary non-moral words generally agree that the same foundational semantic theory applies to moral words as to non-moral words. Still, even though these theorists advocate the same kind of theory for moral words and non-moral words alike, the foundational semantic theories that they prefer are often linked to their metaphysical views. Some philosophers have foundational semantic theories which make it hard to see how any words could refer to non-natural or irreducible moral properties, which leads some of them to believe that moral properties can’t be non-natural or irreducible. In contrast, some other philosophers take this connection in reverse; because they believe that moral properties are non-natural or irreducible, they reject foundational semantic theories which make it hard to see how any words could refer to non-natural or irreducible moral properties.

In his 1988 paper, ‘How to Be a Moral Realist’, Richard Boyd set out the framework for one influential foundational moral semantic theory. Boyd’s theory is a causal regulation theory, and seeks to generalize on causal theories of reference in general by claiming that moral terms, like other natural kind terms, refer to properties, and that the properties to which they refer are the ones to which they stand in a certain sort of causal relation. On Boyd’s view, we do not need a different, special, theory about in virtue of what moral words mean what they do; we only need to apply the best theory that we have about in virtue of what words mean what they do, in general. And he offers this theory in order to make room for the view that moral words do have the same kind of meaning as other, non-moral, words, and in particular that they pick out natural properties in the world.

An important competitor to causal theories like Boyd’s (sometimes known as ‘Cornell’ realism, after Boyd’s institutional affiliation and that of some of his sympathetic colleagues) is the ‘Canberra plan’ espoused by Frank Jackson (so called for Jackson’s institutional affiliation and that of some of his sympathetic colleagues). Whereas Boyd’s theory appeals to causal relations in order to explain how moral terms come to refer directly to properties, and does not make any assumptions about linguistic competence, Canberra planners like Jackson suppose that each moral term is associated with a set of ‘platitudes’ which competent speakers understand to be true. For example, competent users of the term ‘water’ will know that it is the watery stuff of our acquaintance which flows in the rivers and falls from the sky as rain, and similarly, competent users of the word ‘wrong’ will know that when someone thinks that something is wrong, she will tend to be motivated not to do it, unless she is being irrational. Or they might know that if one action is wrong and another is not, then the former action is worse than the latter. According to
Jackson's theory, 'wrong' refers to that property, whatever it is, which best fits the platitudes with which it is associated. Like Boyd's theory, Jackson's theory treats moral words on a par with non-moral words, holding that they require no special foundational semantic theory. But unlike Boyd, Jackson not only seeks to make room for the view that moral terms refer to natural properties, but uses his foundational semantic theory to actually argue that moral terms must refer to natural properties.

In a series of overlapping papers, Terry Horgan and Mark Timmons have argued that any foundational semantic theory whatsoever which explains how moral words come to refer to properties must fail. According to Horgan and Timmons, their view dispatches theories like Jackson's and theories like Boyd's with equal ease. Their argument trades on trying to raise a disanalogy between words like 'good' and words like 'water', on which both Boyd and Jackson model their theories. Whereas Boyd and Jackson assume that 'water' and 'good' are similar, Horgan and Timmons' argument is designed to suggest that they require a very different foundational semantic theory (and consequently, they infer, a very different descriptive semantic theory).

The disanalogy is supposed to be that whereas it is easy to imagine 'Twin Earth' examples, in which people live in a world that is mostly indistinguishable from ours, and speak a language that is mostly indistinguishable from English, but whose word 'water' means something different from ours, it is very difficult to imagine 'Moral Twin Earth' examples, in which people live in a world that is mostly indistinguishable from ours, and speak a language that is mostly indistinguishable from ours, but whose word 'good' means something different from ours. Horgan and Timmons contend that because Boyd and Jackson's theories treat 'water' and 'good' in fundamentally the same way, and since it is possible for there to be a Twin Earth scenario in which there is a world like ours but where 'water' means something different, Boyd and Jackson are committed to holding that it is possible for there to be Moral Twin Earth scenarios in which there is a world like ours but where 'good' means something different. Against this, Horgan and Timmons allege that it is intuitively clear that in any world that is very much like ours in terms of people's behavior, linguistic and otherwise, people's word 'good' would have the same meaning as our word 'good'.

At the heart of Horgan and Timmons' argument is the idea that attitudes matter more than properties to moral disagreement – following an old argument of Hare's that has also been developed by James Dreier, they invite us to consider a world in which people call different things 'good' and 'bad' than we do, but are still generally motivated to pursue what they call 'good' and avoid what they call 'bad'. Each of these authors invites us to infer that despite the difference in what people apply the words 'good' and
‘bad’ to, these people still mean the same thing as we do by those terms. This is what makes Horgan and Timmons believe that moral words are not ‘Twin Earth-able’.

There are several problems with Horgan and Timmons’ argument, one of which is illustrated by inferential role foundational semantic theories, such as that defended by Ralph Wedgwood. According to inferential role theories, words mean what they do because of the inferential or ‘conceptual’ role that those words play. The best examples for such theories are logical words like ‘and’. If a society has a word ‘∗’ with the features that when ‘P’ and ‘Q’ are sentences, people are willing to infer ‘P ∗ Q’ from ‘P’ and ‘Q’, and are willing to infer either of ‘P’ or ‘Q’ from ‘P ∗ Q’, then that is pretty much what it takes for ‘∗’ to mean what ‘and’ does. Wedgwood’s version of inferential role semantics is somewhat complicated, because it is primarily a theory about how concepts come to refer to properties, and leaves out the story of in virtue of what words express the concepts that they do. But the basic idea is that just as ‘and’ is associated with elimination rules as well as introduction rules, the inferential role of ‘ought’ is primarily its elimination rule: that if you believe ‘I ought to do A’, then you are to form not another belief, but an intention to do A. According to Wedgwood’s theory, this is what makes ‘ought’ refer to the property that it does. Wedgwood’s conceptual role theory, or one like it, can intuitively explain why moral terms aren’t ‘Twin Earth-able’; if people were not motivated in accordance with their ‘ought’ judgments, then they would not be using ‘ought’ with the inferential role that gives it the meaning that it has.

2 the nature of moral semantics
As just discussed, Wedgwood accepts a truth-conditional descriptive moral semantics, but uses an inferential-role foundational theory in order to explain why the meaning of moral words like ‘ought’ is closely tied to motivation. In contrast, Horgan and Timmons advocate following Stevenson, Hare, and other philosophers in the ‘noncognitivist’ tradition, and accepting the view that moral words require a different kind of descriptive semantic theory, because they have a different kind of meaning from ordinary non-moral words.

A.J. Ayer articulated a clear version of this idea in Language, Truth, and Logic, where he suggested that ‘he acted wrongly’ does not say anything more, or different, from ‘he acted’. The word ‘wrongly’, Ayer held, does not contribute to what is said by the sentence, but only reflects how it is said – with disapproval. In that way, the meaning of ‘wrongly’ is much like that of ‘the heck’, as in ‘what the heck was Ayer’s theory?’ In the 1950’s R.M. H are offered a more nuanced and sophisticated cousin to Ayer’s theory; rather than saying that moral words don’t contribute to the ‘significance’ of sentences, as Ayer did, H are held that
moral words contribute in the kind of way that mood-markers for imperative and interrogative mood contribute - rather than in the kind of way that non-moral adjectives contribute.

Both Ayer’s and Hare’s theories were therefore built on analogies with other words. But these analogies rapidly break down when we consider more interesting and complex sentences. For example, consider the sentence, ‘if being friendly is wrong, then my parents lied to me’. Ayer’s view, according to which ‘wrong’ doesn’t add to the significance of the sentence, should let us remove ‘is wrong’ to get a sentence that says the same thing, but merely in a different way: ‘if being friendly, then my parents lied to me’. But this sentence doesn’t even make any sense, much less is it plausible that it says the same thing in a different way. To take another example, ‘Max did everything wrong’ turns into ‘Max did everything’, when we take out ‘wrong’. In this case the new sentence makes sense, but does not need to be true, in order for the first sentence to be acceptable. So these are certainly not different ways of saying the same thing.

Some complex sentences make sense on Hare’s theory. For example, Hare carefully noted that we can issue negative imperatives, such as ‘don’t shut the door’, and that these conflict with their positive counterparts (‘shut the door’) in something very analogous to the way that ‘I shut the door’ and ‘I did not shut the door’ conflict. So negative imperatives are a model for negated moral sentences. Similarly, we can issue conjunctive imperatives, like ‘shut the door and open the window’, which, like conjunctive indicative sentences, conflict with the negatives of each of their conjuncts - i.e., with ‘don’t shut the door’ and with ‘don’t open the window’.

But it is hard to mix indicative and imperative sentences under conjunctions or disjunctions: try ‘shut the door and I did it yesterday’ or ‘I shut the door or open the window’. If mixed-mood sentences like these don’t really make sense, that is a problem for Hare’s view, according to which the former is like ‘stealing is wrong and I stole’ and the latter is like ‘I stole or lying is wrong’. Both of these sentences make perfect sense. Ordinary truth-conditional theories of meaning, which treat moral predicates like ‘good’ and ‘wrong’ in exactly the same way as they treat non-moral predicates like ‘tall’ and ‘square’, can easily explain what these complex sentences mean. But if we take Hare’s analogy at face value, it is hard to do so.

Later theorists inspired by Ayer, Hare, and others have proposed that to understand what moral words mean, what we need to do, is not to understand what they are used to do, as Ayer and Hare held, but rather to understand what kind of thoughts they express. In general, according to the central idea of expressivism, to know what ‘P’ means, we need to know what it is to think that P. Moral words and non-moral words have a different kind of descriptive meaning, according to expressivists, because thinking that stealing is wrong is a different kind of mental state from thinking that grass is green. Thinking that grass is
green is just having an ordinary descriptive belief about grass, but thinking that murder is wrong is not having the same kind of belief about murder — rather, it is a matter of having a negative attitude toward murder — call it ‘disapproval’. Because disapproval isn’t just ordinary descriptive belief, ‘wrong’ has a different kind of meaning from ‘green’.

It is well known that expressivists still face a serious problem accounting for the meanings of complex sentences. Following Geach, much of the literature focuses on conditional sentences, but the simplest case is negation. Given the meaning of ‘stealing is wrong’, an adequate compositional descriptive semantic theory should be able to generate the meaning of ‘stealing is not wrong’. Since according to expressivism the meaning of ‘P’ consists in what it would be to think that P, expressivists need a recipe which tells us what it is to think that stealing is not wrong, given what it is to think that stealing is wrong, as input. But no such recipe has been forthcoming from expressivists, and a simple structural assumption makes it hard to see how there could be one.

Nicholas Unwin illustrated the structural problem by asking us to consider the following four sentences:

\[
w \quad \text{Jon thinks that stealing is wrong.} \\
\text{n1} \quad \text{Jon does not think that stealing is wrong.} \\
\text{n2} \quad \text{Jon thinks that stealing is not wrong.} \\
\text{n3} \quad \text{Jon thinks that not stealing is wrong.}
\]

The project of saying what ‘stealing is not wrong’ means, in expressivist terms, is the project of saying what it takes for n2 to be true. But since expressivists have already told us what it takes for w to be true, we can make the following observations:

\[
w \quad \text{Jon disapproves of stealing.} \\
\text{n1} \quad \text{Jon does not disapprove of stealing.} \\
\text{n2} \quad \text{??} \\
\text{n3} \quad \text{Jon disapproves of not stealing.}
\]

The structural problem is that ‘Jon disapproves of stealing’ has less structure than ‘Jon thinks that stealing is wrong’, and so it admits of one fewer place in which we can put a ‘not’. But the place that it is missing, is precisely the place that we need to put a ‘not’ in order to understand what it would take for n2 to be true — which is what it takes to have an expressivist account of the meaning of ‘stealing is not wrong’. This basic structural problem remains at the bottom of a wide range of unsolved problems facing the expressivist
Some new descriptive moral semantic theories depart both from ordinary truth-conditional semantics and from pure expressivism, and advocate the view that moral words like ‘good’ need to be associated with ordinary truth-conditions and with speech-acts or attitudes. (Many of the original noncognitivist theories, particularly including Stevenson’s, were antecedents of this idea.) These theorists suppose that since they are allowing that moral words have ordinary truth-conditions, they will not face the special problems faced by expressivists in accounting for the meanings of complex sentences, but that since they are allowing that there is more to the meaning of moral words than truth-conditions, they can also explain the apparently distinctive features of moral language, including their putative connection to motivation.

Whether such ‘hybrid’ descriptive moral semantic theories will get the results they promise, however, without incurring significant costs of their own, is still an open question; it is already clear that there are many, and deeply different, ways of developing a hybrid theory, and that these ways will differ widely in their prospects and commitments.

3 theory-neutral questions

Even once we settle on a theory about what kind of descriptive semantic theory we need in order to adequately account for the meaning of moral words, and what kind of foundational semantic theory we need in order to explain in virtue of what they mean what they do, there are still a wide range of questions left open about the meanings of individual words. The most interesting of these questions are ones that confront us no matter what kind of foundational semantic theory we accept, and no matter whether we have a truth-conditional descriptive semantic theory, or otherwise. Here we'll be able to consider just a couple of examples.

According to some philosophers, moral sentences are not true or false simpliciter, but require completion in some way in order to be evaluated for truth – by something like a point of view, or a value system. This kind of relativistic view involves a hypothesis about each and every moral term assumed to be so relativized – ‘good’, ‘right’, ‘ought’, and so on – that its semantics involves a relational structure with an extra argument place to be filled by a point of view. So relativism like this is just one example of many issues that can turn on the argument structure of the semantics for moral words.

Another example comes from an ongoing debate about the semantic structure of ‘good’. G.E. Moore famously claimed that ‘good’ denotes a simple and unanalyzable property, in the course of
elaborating his metaethical and normative theory. Against Moore, Geach argued that ‘good’ can’t denote a property, because it is really a predicate modifier. Rather than having only one argument place, as Moore assumed, the word ‘good’ is properly used in sentences like ‘he is a good man’ or ‘he is a good skier’. These sentences are not equivalent to conjunctions of the form, ‘he is good and he is a man’ or ‘he is good and he is a skier’; rather, ‘man’ and ‘skier’ play a role in the semantics for ‘good’.

Another problem with Moore’s idea that ‘good’ denotes a simple property, is that it gave him trouble in making sense of what it means for something to be good for someone. Ethical egoism is the view that each person ought to do what will bring about the most of what is good for her, and Moore famously tried to refute this view by invoking his assumption that ‘good’ denotes a property, rather than a relation to individuals. Given Moore’s assumption, he assumed that egoists had to mean that each person ought to do what will bring about the most of what is good and in her possession, or the most of what it is good that she possesses. And then Moore showed that given either of these two interpretations, ethical egoism is incoherent. But most observers have noted that the tables can just as well be flipped – since ethical egoism certainly makes perfect sense, Moore has a problem, since he can’t make sense of what it is for something to be good for someone. Capitalizing on problems like this for Moore, Judith Jarvis Thomson has argued that ‘good’ never denotes a property, but that it is central to consequentialism that it does, and hence that consequentialism is deeply flawed.

Another kind of theory requires an extra argument place for ‘good’ in order to explain agent-centered constraints in a consequentialist framework. If we understand consequentialism as the thesis that everyone ought always to do what will bring about the most good, it appears to be inconsistent with what are known as agent-centered constraints – things that you ought not to do, even if by doing them you can get more other people not to do them. For example, it is plausible that that you ought not to murder, even if by murdering you can prevent two other murders (perhaps it is okay to murder to prevent a thousand murders – agent-centered constraints need not be absolute side-constraints). Ordinary consequentialism can’t allow this, because there is no way of ranking how bad each murder is, which can explain why each person ought not to murder even in order to prevent two other murders. Consequentialists can explain why one person ought not to murder even in order to prevent two other murders – but only by assuming that the first person’s murders would be worse than the others. But then according to consequentialism, it would make sense for other people to murder in order to prevent the first person from murdering. There is simply no ‘agent-neutral’ way of evaluating how good or bad outcomes are that can account for agent-centered constraints in consequentialist terms.
Some theorists who accept the existence of agent-centered constraints therefore argue that we need to think that ‘good’ has an extra argument place for an agent, so that ‘everyone ought always to do what will bring about the most good’ can be interpreted as ‘∀x(x ought always to do what will bring about the most good,)’. If ‘good’ is agent-relative in this way, then even though there is no single way of evaluating how good or bad outcomes are that can capture constraints, there may be a combination of ways of evaluating how good or bad outcomes are – namely, one for each agent – which allows us to explain constraints. This leads to a theory that is supposed to be a generalization of consequentialism that can escape some of its major problems. But it turns on assumption about the logical structure of ‘good’ – namely, whether it has a surprising extra argument place for an agent in this way.

All of the issues that I’ve mentioned so far turn on the argument structure of ‘good’, which is a theory-neutral semantic question. Any descriptive semantics for ‘good’ will have to build in an understanding of how many argument places ‘good’ has, what sort of things fill those argument places, and how the things filling those argument places are determined. The same thing goes for expressivists and truth-conditional semanticists alike – although here, too, things can get complicated, as some authors have argued that if ‘good’ has more than one argument place, that will be a problem for expressivists. In the abstract, it’s hard to see why this would be so, but Horgan and Timmons’ expressivist theory gives us an example of how at least one expressivist theory can run into a similar kind of trouble.

Horgan and Timmons’ theory doesn’t tell us anything about what ‘good’ or, for that matter, any other moral adjective means. But they do tell us what ‘ought’ means. On their view, the difference between ‘Max runs’ and ‘Max ought to run’ is that ‘Max runs’ expresses an is-belief in the proposition that Max runs, and ‘Max ought to run’ expresses an ought-belief in the proposition that Max runs. Same proposition, but different attitudes. Theirs is an expressivist view, because it holds that rather than contributing to the truth-conditions of the sentence, or to what someone believes, who accepts a sentence in which it figures, ‘ought’ functions to change the kind of belief that you have, when you accept a sentence in which it figures. Rather than an is-belief, it is an ought-belief.

In order to work, Horgan and Timmons’ view therefore requires the assumption that ‘ought’ has exactly one semantic argument, which is filled by a proposition. This is the kind of picture of ‘ought’ that is manifest in some ways of interpreting Standard Deontic Logic, according to which ‘ought’ is a propositional operator, and all simple sentences involving ‘ought’ amount to saying that something ought to be the case. (The thesis that ‘Max ought to run’ is equivalent to ‘it ought to be that Max runs’ is sometimes known as the Meinong-Chisholm Thesis.)
This theory about the nature of ‘ought’, however, is very controversial. According to John Broome and Ralph Wedgwood, for example, though ‘ought’ is indeed a propositional operator, there is one important sense of ‘ought’ on which it has a separate argument place, for an agent, in addition to its argument place for a proposition. If Broome and Wedgwood’s view is right, then sentences involving that ‘agential’ sense of ‘ought’ cannot be understood in terms of a distinctive kind of attitude that the speaker bears toward a proposition, as Horgan and Timmons claim, because such an account could not distinguish between accepting the ‘ought’ sentence relative to one agent, and accepting it relative to another. Moreover, other theorists, prominently including Peter Geach and Gilbert Harman, have gone farther, arguing that ‘ought’ does not have an argument place for a proposition at all. If Geach and Harman are right, a view like Horgan and Timmons’ would certainly fail.

Extra argument places for ‘ought’ have also been argued for from other directions. According to a standard linguist’s treatment of ‘ought’, which offers a unified explanation of how ‘ought’ works in both deontic and ‘epistemic’ uses, ‘ought’ sentences require both a modal base and an ordering source in order to be semantically interpretable. And another dispute concerns whether and in what sense ‘ought’ is relative to evidence. In poker, it may be that a player ought to fold, relative to her own evidence, but that she ought to call, relative to the evidence of someone who has seen both her cards and her opponent’s. Any of these extra argument places for ‘ought’ would raise problems for Horgan and Timmons’ theory, which is just one example of many of how questions about argument structure, which are theory-neutral in nature, can have broader implications for other questions about moral semantics, as well as about broader questions in both normative and meta-ethics, as well.
references


Geach, Peter [1956]. ‘Good and Evil.’ Analysis 17: 33-42.


