The Dimensions of Moral Semantics (unpublishable draft)

David Brax, Deparment of Philosophy, Lund University
david@brax.nu

Introduction

The development of two-dimensional semantics has been put to great uses in different areas of philosophy. Particular in the philosophy of mind (see Chalmers 1996). I recently became interested in how it would do when applied to moral philosophy. What we want is a way to characterise moral claims semantically, we want to account for how moral concepts relates to natural concepts and properties.

The approach seemed promising and, as some writers have recognised (Gibbard, Dreier), we might want, indeed we might *need*, to extend the technique and give three dimensional semantic analyses of normative concepts.

In this paper I account for these attempts and how I believe they fare. I begin with a brief story about two-dimensionalism. I then go on to see how it might be applied to moral semantics. Finally I account for two three-dimensional attempts, and argue for one of my own. (It is a coincidence that part III that deals with three-dimensionalism has three different accounts in it, but I guess Kant would have been proud.)

Part I
Why two-dimensionalism?
The two-dimensional approach to semantics associates linguistic items (expressions, utterances) with two different sorts of semantic values, were "semantic values" can be understood as *meanings* or *contents*. The first of these semantic values gives the reference and truth-conditions for the utterance; the other gives how these two features depend on the context of the utterance. The point of this second dimension is to enable analyzing matters of *cognitive significance* and context-dependence, whereas the first dimension gives the metaphysical, modal status of the utterance. What do we want with the second dimension for? To be specific: we want it for being able to say that even though water is H2O as a matter of metaphysical necessity, it might have turned out differently. We want it to account for the epistemic part (as distinct from the modal) of semantics. Meanings, after all, are present not just to determine what is true, but what *might* be true given what we know, what is left open even with linguistic competent attribution of predicates, uses of sentences. Two dimensional semantics gives the relationship between these two notions via a two-dimensional matrix. The best way to ease into it is by a case-study.

a) The butler did it: a case study
A murder has been committed and the detective makes the bold claim (S).

(S) The butler is the murderer.

 Let's assume that this is a true claim, and let's further assume that 'the butler' and 'the murderer' identifies uniquely. How would we characterize this claim semantically?
Determining the truth of (S) is certainly an a posteriori matter: The revelation that the butler is the murderer should be answered with a gasp, not with accusing the investigator for stating the obvious (it is not trivial in a technical sense, even though it might be trivial as a plot-device). By contrast, claiming that the butler is a *servant* seems to be necessarily true.
On an extreme version of Kripkean semantics, (S) is necessary whereas 'the butler is a servant' is contingent. The butler is certainly not the murderer by *definition*; the butler is the murderer because the two phrases identify the same person. On this kind of reading, the butler is not a servant necessarily; the butler is not even a butler necessarily.
Metaphysically, since the butler and the murderer is the same person, (S) is necessarily true. A person is necessarily self-identical. By contrast, it is not necessarily true that the butler committed the murder. This means that 'the butler' is 'the murderer' even in worlds were the person picked out is neither a butler, nor guilty of committing the murder. In those worlds, the (still true) claim (S) is not only not informative; it is positively misleading.
This is an awkward way of characterising (S) semantically, but only marginally more so than the familiar Kripkean claim that 'water' refers to 'H2O' even in worlds were H2O is not a watery substance, and something else, like XYZ, is. It is certainly part of "folk-semantics" that if someone is "the murderer" that someone must have committed a murder, but, admittedly to a somewhat lesser extent, it is part of "folk semantics" that if something is water, it has most of the characteristics of water. Some might want to claim that while water is necessarily a liquid at room-temperature (etc.), it is not necessarily H2O. Kripkean semantics gives us the complete opposite of what we want in these cases.
When it comes to "murderer" and "butler", the awkwardness of the semantic characterisation above because it disconnects the terms from the application of the associated descriptions. The claim (S), we are likely to claim, is equivalent to

(S*) the butler committed the murder

In fact, (S) is *cognitively equivalent* to (S*), but they do not, on our reading, have the same modal status.

   Clearly we have two ways of talking about necessity that does not seem incompatible with each other. When the terms are treated as proper names, (S) is necessary. When either of them, or both, is treated as a description (S) is contingent. Both seem to be relevant for the question of the meaning of (S). Semantics ought to be able to characterize both claims, and that is precisely what two-dimensionalism does.

b) Two-dimensionalism: Let's get technical
Extensions of sentences is their truth value, of a singular term (names) it is the referent, and for general terms it is the class of individuals that fall under the term. Intensions, on the other hand, are functions: For a sentence, it is a function true at a possible world if the sentence is true there, for a singular term it is a mapping of a possible world to the referent of a term in that world, and for a general term maps a possible world to the class of individuals that fall under the term in that world (Chalmers, forthcoming).

   Two expressions that have the same extension can differ in intension. Two general terms can refer to the same class in this world, but differ in other possible worlds. Usually, this difference is due to a difference in *meaning*. Quine's case of the co-referring general terms 'cordate' and 'renate' is a case in point. Our murderous butler is another. In many possible worlds, these actually co-referring terms does not co-refer. Carnap noticed that intensions behave a bit like Fregean senses: intensions gives the aspect of an expression's meaning that corresponds to its cognitive significance. When utterances are cognitively equivalent, their extension will coincide in all possible worlds, and hence they will have the same intension.
Kripke claimed otherwise: Many statements that we know to be true only empirically are true in all possible worlds. Water is H2O, Kripke claimed, in *all* possible worlds, and hence have the same intension, even though they differ in cognitive role.

 "On the face of it", Chalmers writes, "cognitive differences between the terms is connected in some fashion to the existence of these possibilities [that the terms would fail to co-refer]. So it is natural to continue to use an analysis in terms of possibility and necessity to capture aspects of these cognitive differences." (p 2) There is a sense in which for a term like 'water', the term's extension and its Kripkean intension depends on the character of our world. That is, if some other world (say Putnams twin earth) would have turned out to be actual, water might have been XYZ. The (Kripkean) intension seems to depend on the character of the world. This dependence, on two-dimensional semantics, can be represented as a function from worlds to intensions. (Kripkean intensions are functions from worlds to extensions). Diagramatically, we can represent it like this:

	H2O-world	XYZ-world
H2O-world	H2O	H2O
XYZ-world	XYZ	XYZ

This diagram expresses an aspect of the term 'water'. It shows that if the H2O world is actual (i.e. is the world of the utterance), H2O is picked out by the Kripkean intension for 'water' in all evaluated worlds. If, on the other hand, the XYZ world were actual, 'water' would pick out XYZ in every possible world. Roughly, we could say that the column on the left picks out epistemically possible contexts[1] (if you happen to be in the H2O context, your term 'water'

---

[1] For the sake of simplicity, contexts in this case are worlds.

will pick out H2O), and the row on top picks out metaphysically possible worlds. The column lists worlds *considered as actual* and the first row lists worlds *considered as counterfactual*. In some cases, more precisely in the cases on the *diagonal*, the world considered as actual and the world considered as counterfactual will coincide. The diagonal gives how a term's extension depends on the context of utterance. We could call this the terms *diagonal intension*. The matrix as a whole gives a *two-dimensional intension*. What's important here is that terms with the same extension ('cordate' and 'renate') and terms with the same Kripkean intension ('water' and 'H2O') can have differing two-dimensional and diagonal intensions. So how does this matrix relate to cognitive significance?

The Chalmers' version of two-dimensionalism connects the first-dimensional semantic values to a priority and cognitive significance in a strong way. Not just for indexicals and descriptive names (As was done by Kaplan and Evans), but for expressions of all kinds. At least one intension of an expression is strongly tied to the role of the expression in reasoning and thought. Chalmers summarizes the core claims in the following manner:

1) A primary intension of an expression token is a function from scenarios to extensions. A secondary intension is a function from possible worlds to extensions.
2) A sentence token S is metaphysically necessary iff the secondary intension of S is true at all worlds
3) A sentence token S is a priori iff the primary intension of S is true at all scenarios.

A scenario is a centered world. The secondary intension is the intension that picks out the extension of the expression in counterfactual worlds. 'Water' picking out H2O, 'I' picking out me, etc. An expression's primary intension for 'water' is, very roughly, the clear, drinkable fluid with which the individual at the center of the scenario is acquainted. The primary intension coincides with the diagonal of the two-dimensional intension (i.e. the world considered as actual both as scenario and evaluated world). The secondary intension corresponds to the row beginning with the world considered as actual. In addition to the three claims above, we get:

4) A sentence token S is necessary a posteriori iff the secondary intension of S is true at all worlds but the primary intension of S is false at some scenario.

5) A sentence token S is contingent a priori iff the primary intension of S is true at all scenarios but the secondary intension of S is false at some world.

The necessary a posteriori sentences are those that have a necessary secondary intension, but a contingent primary intension. The contingent a priori has a contingent primary intension but a necessary secondary intension. Primary intensions, then, behave much like Fregean sense behaves. Chalmers define them as *epistemic possibilities*. Utterances have two intensions: they are associated with two *propositions*, we might say. Which one we "intend" is revealed by how we evaluate possible, non-actual worlds. There is no way of saying which is *the* intension or *the* proposition.[2]

---

[2] In the Butler/Murderer case, we have a "the butler did it" world (B+) and a "the butler didn't do it" world (B-) and get, for (S):   and for (S*)

	B+	B-		B+	B-
B+	T	T	B+	T	F
B-	F	F	B-	F	F

Here, clearly, the secondary (Kripkean) intension (B+ row) gives the less plausible characteristic, while the primary intension (diagonal) gives the plausible one. Of course, some would want to say that (S) evaluated for the B- world should give an F. This is so on evaluation of (S*):

We'll now turn to the application of this framework for moral semantics.

Part II
Two-dimensional moral semantics: Intensions for all intents and purposes

Can two-dimensionalism be applied in an enlightening way to the field of moral semantics? What is it that we want with moral semantics? We want to be able to show how truth-values are to be distributed to moral utterances and we want to account for the "normativeness" of these utterances. Exactly how this should be done is of course substantial ethical and meta-ethical questions: what we want with moral *semantics* is a way of characterising these substantial ethical and meta-ethical claims. Two-dimensional could perhaps provide the framework that we need.

As we have seen for natural kind concepts like 'water', we can construe two-dimensional matrices that give some aspect of the meaning of and distributing truth-values for assertions employing them. Could we do the same for *normative* concepts? If we are (one type of) naturalists or realists, moral concepts will be just like natural concepts and hence should behave just like they do.

Gibbard (2003) argues that we all are *committed* to descriptivism, or at least to some extent. He argues, and I will here accept without repeating the argument, that we are committed to the claim that actions are right in virtue of non-moral properties. That is, if an action is right, it is right in virtue of some non-normative properties that it has. Every plausible moral theory makes a claim of this sort.

Fixing reference or giving meaning?
How do we account for claims like (R): "Action X is *right*", and how do they relate to claims like (H): "Being right is being egohedonic"?

If it is true that being right is being egohedonic, action x is right only if action x is egohedonic. If 'being right' and 'being egohedonic' are identical, settling that the action is egohedonic is all that it takes to establish that it is right. But, if Moore's open argument shows anything, being right and being egohedonic can not have the same meaning[3]: The concepts differ in cognitive significance. Moral utterances may very well pick out a natural class (say the egohedonic one) but their *meaning* seems to be different. We can agree about whether an action is egohedonic, but disagree about whether it is right. Naturalists identify *properties*, not necessarily *concepts*[4]. As we have just seen how two-dimensionalism accounts for differences in cognitive significance for (necessarily) co-referring terms in terms of diagonal intensions, we might be optimistic about the same strategy as applied to moral semantics.

If egohedonism is correct an action will be right if and only if it is the egohedonic action. How should we characterize this in two-dimensional semantics?
Here's a start: the egohedonist claims that an action x will be right in precisely those cases were x is egohedonic. Let's say that in W1, X is egohedonic, and that in W2 it is not, we get

For "X is egohedonic":   and for "X is right":

The difference is due to that "committed the murder" designates a property, not an individual. As it happens, 'committed the murderer' is the primary intension of 'is the murderer', hence the row for B+ for (S*) will be the diagonal for (S). Similarly for "water" and "watery substance of our acquaintance".

[3] Even if they could be referring to the same property

[4] Naturalist egohedonists *could* say the following in response to the open question argument: Some one who realise that an action is egohedonic but failed to see that it is right would simply lack knowledge about rightness.

	W1	W2			W1	W2
W1	T	F		W1	T	F
W2	T	F		W2	T	F

The two claims have the same two dimensional intension and (hence) the same diagonal (i.e. primary) intension. Here, nothing depends on what world is considered as actual. By contrast, the two-dimensional matrices of "X is water" and "X is H2O" differ from each other. The difference is that, by stipulation, neither 'is right' nor 'is egohedonic' here refers rigidly: X varies from world to world in both respects. By the looks of it, then, two-dimensional semantics cannot account for the difference in cognitive significance between 'x is egohedonic' and 'x is right' in the way it does with water and H2O. A two-dimensional matrix for the identity claim (H) would be filled with T's:

(H)

	W1	W2
W1	T	T
W2	T	T

(H) on this reading is necessary and a priori. Is this the gist of the identity claim made by egohedonists? Do they want to say that the identity is a priori? If this is compatible with the open question argument, we must abandon the claim that this two-dimensional matrix can answer questions on cognitive significance. But what if the identity claim is more of the Water = H2O kind?

Necessary a posteriori
Do moral terms refer rigidly?[5] Ego-hedonists could be understood as making the claim that "being right" refers rigidly to the property of being ego-hedonic, i.e. that it picks out the ego-hedonic acts in all possible worlds. Most interestingly for our purposes, it would then pick out the ego-hedonic acts even in worlds were the ego-hedonic lack the motivational force it has in our world. The property of being ego-hedonic here is not picked out merely by being ego-hedonic, but by being *right*. It would, of course, be desirable to know what the cognitive significance of 'being right' is, we could then see how egohedonic act could be the acts referred to. Actually, we do know *something* about how 'being right' works that is not dependent on our having the correct moral theory. That is why moral theories can be characterised as *conflicting*, after all. Jackson (1998) suggests that the property of being right is the property that plays "the rightness role". He says two things about it. First: the "rightness role" could be described in non-moral terms, and second: we don't need to know what our concept 'right' refers to in order for our uses of the term to refer successfully.
The open question argument seems to establish that the cognitive significance of 'being right' is not identical to the cognitive significance of that which is right (i.e. egohedonic).

Two-dimensionalism enables us to say the following: If there is a 'rightness role', (H) could be construed as the claim that in our world, that role is played by the ego-hedonic. But if we consider some alternative world as actual, whatever plays the rightness-role there will be what makes rightness claims true when uttered there. "Playing the rightness role" would be the *primary intension* of 'being right'. 'Being right' would then refer rigidly from worlds considered as actual. The diagonal intension of (H) on this reading would not render ego-hedonism true throughout the matrix. We would than have what we could call "Cornell

---

[5] Jackson (1998) raised this question but did not make a commitment to either (even though he believed that folk-morality tends to accept rigid designation for moral terms).

hedonism". If you are an ego-hedonist of this sort, your claim "Being right is being egohedonic" will be a matter of reference, not "meaning", and it will be necessary a posterori.

If what we have said here is correct, it might have turned out that the rightness role was played by something else, just as it might have turned out that the watery stuff was XYZ. Jackson wants to remain neutral on the epistemic status of moral claims (whether they are a posteriori or a priori). He talks about it "turning out" that the rightness role is played by hedonic states, which suggest that he would accept a diagonal intension were, if things had turned out otherwise, the moral properties would be the one they then had turned out as being. (The Cornell claim of a posteriori identity of properties, without analysis, though (this is precisely what Gibbard fixes with his plan-laden dimension)) This neutrality, though, concerns whether moral identity claims are claims about reference fixing or about meaning. If we allow for one dimension for each, we can see what the discussion is about, even if it doesn't help us settle anything. There might be no such thing as *the* intension of such ethical claims, but in meta-ethical debate you could characterize your statements by how you would go about evaluating your claim for possible worlds considered as actual and counterfactual. A weaker form of realism (not Cornell) would make the reference a-non rigid matter: see below.

If what has been said is true the claim (T) "The right is what plays the rightness role" is contingent but a priori: it will be true on the diagonal. This diagonal indeed is equivalent to the cognitive significance of 'being right'. Here, this has been a matter of stipulation.

What *is*, roughly, the cognitive significance of being right? In order for conflicting moral views to be discussing *the same thing*, there must be some intension they share. In the debate whether H2O or XYZ is water, we are talking about the same thing because we are talking about the watery stuff of our acquaintance. Something similar must be the case for moral debate to be proper debate.

In *our* world, being egohedonic is what plays the rightness role, but in W2, something else does. Let's say that in W2, people have a hedonically inversed spectrum the rightness role is played by what is ego-*doloric*, i.e. of what gives the most pain. Say that X is the egohedonic action, we get the following for "X is right", (R):

(R)

	W1	W2
W1	T	T
W2	F	F

This differs from the matrix for "x is ego-hedonic" above, which explains why the question is open. If we accept that "playing the rightness role" roughly approximates what we mean with "being right" the cognitive significance of "being right" will be given by the diagonal. If, on the other hand, we are set on what is *actually* right, our intension will be roughly equivalent to the row from W1: being right is being ego-hedonic, since that is what gives reason in *our* world.

You *might* want to characterize the identity claim in these two manners:

(H)

	W1	W2		W1	W2
W1	T	F	W1	T	T
W2	T	F	W2	F	F

In the first case, 'right' does *not* refer rigidly (it is not like water-H2O), whereas in the one to the right it does. These are two quite different forms of egohedonism. Generally: As long as

there is a T in the upper left corner, you will be (some kind of) hedonist, as long as there is a T somewhere in the matrix, you will agree that Hedonism could have (and perhaps has) turned out to be true. If there are *no* T's, you will be a rabid anti-hedonist. If there is nothing but T's in it, you will be a hedonist-extremist, insisting that worlds were the rightness role is played by something else is a world gone mad. It is probable that such a hedonist would accuse the above for not being properly hedonist at all; that it lacks commitment to the hedonist cause.

Remember that in the "water is H2O" case, the identity claim gets an F when your standpoint is a world in which the watery stuff with which the person at the center is acquainted is not H2O. The parallel case would be the identity claim (H) from a world in which the right action is not the egohedonic one. That would be a world in which ego-hedonism is *false*. Naturalist Hedonists typically want to say that there is *no* such world, whereas non-hedonists typically will want to say that *all* worlds are such that the identity claim is false (there might be worlds where the egohedonic actions are the only right ones even on a non-hedonist reading, though).[6] Actually, the extremist hedonist can accept that *as said by the people in that world*, hedonism is indeed false, but that is because 'being right' do not mean what it means here, it is not even the same concept. If *our* use of the term is what gives the true meaning, and hence if we want to know what is right to do in another world, we should look how it is evaluated when *our* world is considered as actual and theirs as counterfactual. This view will then be like our hedonist-extremism with the difference that it will be committed not to hedonism per se but to whatever turns out to play the rightness role in the actual world. It would be a sort of semantic chauvinism.

Here is what seems like a problem for the two-dimensional characteristic given: *If you accept one of the intensions in the twodimensional framework for (H) as necessary a posteriori as *the* intension that gives the meaning of the term 'right', you might insist on a two-dimensional characterisation of your claim that differs from that matrix. As we have just seen, the die-hard hedonist wants a matrix full of T's for (H), the "rightness role" sympathizer could aspire to be world-sensitive and hence accept the evaluations of the people in the world considered as actual, and hence want a matrix for (H) with T's only in and for worlds were the rightness role is played by the egohedonic. These two possibilities are for cases where the claim (H) is *not* intended as necessary a posteriori. The hedonist-extremist will have difficulties with the open question argument, but might adopt the strategy of footnote 2. The lighthearted hedonist will have difficulties with the supervenience thesis. This means that the matrix suggested by Jackson's way of putting matters and the Cornell realists is not meta-ethically neutral.

Moral Twin Earth is Calling for You to Decide: When in twin-Rome…
If someone on twin-earth, fantastically, would call you and ask for water, you probably would recommend XYZ. It is, after all, what he asked for. If, on the other hand, even more

---

[6] Some hedonists might agree, and say that hedonism in fact is contingently true: i.e. there is a class of right actions, and in some worlds this class is the egohedonic class. An extremely light-hearted version of hedonism would say that it just so happens that our world is such. A little less light-hearted would be the claim that being ego-hedonic is what *makes* acts right in our world, but it could have turned out differently. Again, a little tougher this time, as it *happens* that being egohedonic is what makes acts right, it does so in all world (actualism). And the toughest version, as we have seen, is the one were no matter were you are and what you consider, only ego-hedonic actions are right. If we allowed that right in some world would not be ego-hedonic, we could say that it is not *really* right in the same sense that we can say that XYZ is not *really* water. The versions of ego-hedonism here could be characterized by the status they assign to (H). They will all agree that it is true, but they will say that it is contingent a posteriori, necessary a posteriori, necessary apriori or contingent apriori (check on it)

fantastically, you are to do the right thing on moral twin earth (were ego-hedonism, somehow, holds), what should you do? The problem is that there is no one answer to that question. You'll have to decide which is the relevant intension of the term 'right'. You have to decide whether it is the context of your utterance or the evaluated world that is is the 'focus of the utterance', so to speak. And were to put the emphasis is a matter of meta-ethical (as distinct from normative) commitment.

What we have seen in this section is that the claim (H) can be characterised in a two-dimensional way, but also that you can characterise it in *different* two-dimensional ways. This means that (H) is not exhausted by any one two-dimensional characteristic. Perhaps a third-dimension is needed.

Part III Three dimensional moral semantics: standpoints of evaluation

In "Thinking how to live", Gibbard makes a case for expressivism for a certain class of concepts he describes as "plan-laden". He starts out carefully talking about "Being the thing to do" as a typical plan-laden concept, but later on he argues that our normative concepts are plan-laden as well. As I said before, Gibbard argues that we as planners are somehow committed to descriptivism. Being the thing to do is dependent on it having (broadly) *natural* properties of some kind or other. If ego-hedonism is right here, the ego-hedonic action is always the thing to do. But whereas calling something "the thing to do" *settles* what to do, calling it "ego-hedonic" doesn't. The argument in "thinking how to live" is that *given* a certain plan, the descriptive properties of actions settles what to do. That is, whereas only natural properties exist, non-natural *concepts* exist that picks out these properties and assign roles to them. The open question argument can only show the distinctness of concepts, not of properties.

We speak, Gibbard notices, *as if* naturalism were true. His theory wants to accommodate this sort of speaking within an expressivistic theory for normative concepts. He does so by appealing to *plans*[78]. Moral concepts *relates* to natural concepts: for every plan there is a (broadly) natural concept that in turn have a two-dimensional matrix, as seen above. But which matrix gives the extension (and primary intension) of "the thing to do" depends on the plan.

The plan is what makes for the difference between "the thing to do" and "being ego-hedonic" (if that is the plan): it accounts for the difference of cognitive significance when characters coincide. The claim (H) is not purely descriptive. Being right here *signifies* being ego-hedonic, but it *expresses* a concept that is distinct from the concept of being ego-hedonic. The property of being ego-hedonic, if (H) is true, *realizes* the concept of being right. To say that being ego-hedonic realizes the property of being right is to settle a moral issue, it consists in accepting the egohedonism (see Gibbard (p115-6)). In comparison:

How should we understand the relation between concept and property. In the case of water=H2O, it is a matter of the *indexicality* of the behaviour of 'water' (and non-indexicality of H2O) that settles that it is H2O. In the matter of "Being the thing to do" and "being egohedonic", they differ, not because one is indexical, but because the one is plan-laden whereas the other one is not.

---

[7] Actually *hyper*plans, plans that determines what to do in every situation and that are such that we could come to accept them without changing our minds, again Gibbard argues that we are *committed* to the "existence" of such hyperplans

[8] "Plan" is better than "the rightness role" insofar as it does away with the questionable issue of defining that role, it has other difficulties, though.

Gibbard argues that concepts have *characters*, and characters are like our two-dimensional matrices. "The character of a concept gives its extension – its truth value, what it designates, or the like – as the concepts is applied from various standpoints (…) to various possible ways things might have been". We can translate "Standpoints" to "worlds considered as actual" (centered) and "possible ways things might have been" to "counterfactual worlds". The character is always natural in the sense that some broadly naturalistic concept might share this character. This means that the character for 'being right' will be *shared* by some natural concept, say being ego-hedonic. But, as we saw in the last chapter, if they share two-dimensional matrix, we cannot keep them apart, and we want to do so *even* if being ego-hedonic is what makes it true that it is the thing to do. Gibbard writes "with plan-laden concepts, at least, the character of the concept doesn't exhaust the concept; distinct concepts can share one and the same character".

As planners we can say that it is *as if* the concept (being right, say) had a natural character.

> "Like any concept, the concept of *being egohedonic* has a character. That also, then will be the character of *being okay to do* – or its quasi-character, if you wish. This just amounts to saying that by *a priori* necessity, the two concepts are coextensional. (…). The character is a function of standpoint-world pairs, a standpoint *s* of thinking paired with a world *w* thought about. (…) take any standpoint *s* and any world *w*. Suppose that *s* is one's standpoint, and from standpoint *s* one is thinking what would be okay to do in world *w*. To the pair [*s,w*] the character of *egohedonic* assigns those acts that would then be egohedonic. To the same pair, the character of *okay to do* assigns those acts that would then be okay to do." (p 127)

Here, stand-points does not include (hyper) plans.

Extended character:
In order to account for the fact that plans determine which natural character is shared by our plan-laden concepts we need to extend the character. The meaning of plan-laden concepts is in part given by answering questions about how to live.
The planning dimension: It assigns truth values to triples of standpoint from which you think, world thought about and (hyper)plan. Gibbard does not say whether these extended characters fully exhaust the content of thoughts (as we shall see, they probably don't).
  Given acceptance of a hyperplan (a way to live), you have a twodimensional character for how to determine it. Does plans identify their natural character a priori? Gibbards say they do, but as we have seen in the former section, this might be questioned. He says "A fundamental answer to the question of how to live would claim *a priori* status". As I have said, we *might* want to do so, but it is certainly not a matter of conceptual necessity. *My* third dimension would therefore *include* this further meta-ethical commitment (or take on a dimension of its own). (I will return to the matter). Our hedonist-extremist would like to make the claim he does.
Two persons can disagree about how to live, but agree about the issue debated by agreeing on extended character. To determine the extension of "right", you must settle how to live, this is captured by the extended character.

Why not settle for two dimensions, by including the plan in the standpoint? The reason is that we want to be able to talk with people with differing ideas, and we want to be able to tell them what to do, even when it conflicts with their (hyper)plans. If you work in the plan in the two-dimensional, you loose explicitness. The third dimensions there to make it possible for us to say that it was wrong for you to do what was right for you to do and vice versa[9]. When it

---

[9] In a similar spirit, you want to say that you can be both right and wrong in saying 'Water is H2O' when visiting twin-earth

comes to morality, we want to say that contingently co-referring theories differ in meaning. If deontology and utilitarianism yields the same results, this is a case in point. (Plans that pick out the same actions in our world but differ in other are different from each other: if the utilitarian and virtue ethical would coincide, for instance, they will not do so for all worlds, that is why we cannot settle for two-dimensions by cutting the possible worlds of evaluation out.)

Gibbard, then, adds the planning dimension to account for (the "internalistic" part, as it were, of) the cognitive significance of plan-laden terms.

[*A semi-detached section about Dreier's three dimensional account for meta-ethics*
James Dreier (2002) has quite different reasons (he wants to be able to express a meta-ethical view that does not commit normatively) for developing a three dimensional semantic network. Were we have spoken about "the rightness role" and Gibbard about "hyperplans", Dreier speaks of *moral standards*.

> The rough idea is that if I assert a statement that is true relative to some moral standards but not others, you will be able to draw some conclusions about my moral standards. I will have committed myself to standing by one or another of those moral standards that count the statement true." (p 253)

He starts two-dimensionally, not considering possible worlds for now, but moral standards instead. As an example, he takes the sentences (F) "Making fun of me is wrong". Let A1 be G.E. Moore, A2 be Ray Charles and A3 be the Pope. Then take three moral standards: M1: making fun of the blind and the catholic is wrong, M2: making fun of philosophers and the catholic is wrong and M3: it is never wrong to make fun of people. We get:
(F)

	M1	M2	M3
A1	F	T	F
A2	T	F	F
A3	T	T	F

The diagonal here is not interesting, since there is no special connection between the speakers and the moral standards. But if we arrange it so that for every A(i), M(i) will be the standard of A(i). For instance: take the claim (W) "X is wrong if and only if it is counted wrong by my moral standard. Let's say that on A1's moral standard M1, X is wrong, and for A2, who have moral standard M2, X is not wrong. We get
(W)

	M1	M2
A1	T	F
A2	F	T

According to M2, A1 has the wrong idea (namely M1) and according to M1, A2 has the wrong idea (namely M2). But the diagonal gives all T's which means that the claim is, as Dreier would have it, *morally non-committal*. This is so because it holds for all X's no matter what your moral standard is. It is always true when the world is considered as both actual and counterfactual.
Now let's consider the addition of possible worlds to this:
We want to be able to talk about *modal* claims, like (C): "Although Capital punishment is wrong, it could have been permissible". Even if Capital punishment could *not* have been permissible, we want to be able to represent it. How? We could, Dreier says, add two

dimensions, one for the possible world as context of utterance and one for the possible world evaluated. But four dimensions is a bit hard to represent, so we could make it into three.

> "We'll think of a moral proposition as delivering a proposition (a set of possible worlds, or an assignment of truth values at each world) given a moral standards. So it is a propositional function, taking moral standards as arguments. Then a moral proposition is itself a matrix with worlds as one dimension and moral standards as the other. And for contexts we'll take not speakers, or worlds, but speakers-at-worlds" (p 255)

A cell in such a three dimensional matrix, then, is specified by a context, a world and a standards. This we could draw as a box, or we could just chicken out and draw it as a set of two-dimensional matrices where one of the dimensions is represented as assigned to each matrix. Each standard, context or world for a moral sentence picks out a *slab* of the three-dimensional matrix and each slab is a two dimensional matrix. The matrix for (W) above, for instance, is the slab where the world of evaluation is fixated as *this* one.

Why not make it even simpler by including the moral standard of the speaker-at-world so that we get speaker-at-world-with-moral-standard on one axis and world evaluated on the other? Again, what we loose is explicitness (but do we loose the possibility to express modal claims? No we don't, as I will say, modal status could be the third dimension. Alright, then, Dreiers version is not more complicated).

As I've said, Dreiers objective in his paper is to (in answer to Dworkin) find non-committal meta-ethical assertions. Such an assertion we will get if:
For a world w* and speaker a*, the moral proposition (which is a two-dimensional "slab") expressed by the sentence at the context a*-at-w* is true at w* according to the moral system held by a* at w*. Moral noncommittal sentences have T's all along their "major diagonal". The major diagonal runs from the corner were the first world, context and standard is, and cuts through all three dimensions to the last world, context, and standard of the matrix. An example would be (A): Necessarily, X is wrong if and only if my actual moral standards count it wrong.
Note that this *implies* that X is wrong iff my actual moral standards count x as wrong, for all X's and all moral standards. Asserting (A) does not commit me to anything in particular. Commitment follows only from what my moral standards really are.
In a *context* it expresses a moral proposition (i.e. the slab given when the context is fixated).]

Does neutrality require four dimensions?
There is some point in characterising ethical (and in Dreiers case: meta-ethical) claims by making place for more dimensions on which they can be characterised. To make explicit how the extension of moral claims depends on context and moral standard (or plan) is part of making sense of those moral claims. What we want is a *neutral* way of characterizing moral claims semantically, i.e. a way that holds no meta-ethical, modal or normative *commitments*.

My ambition as hinted at in the second part differs a bit from Gibbard's. The feature I was missing in the expositions there was that we might want to claim that plan-laden claims could actually *fail* to have the same character as the truth-maker of those claims. These are the cases were 'being right' is *not* 'ego-hedonic', or what not, a priori. The point I wanted to make is this: Gibbard says that he is not sure that identity in extended character is enough for sameness of content. I would say that it isn't, and that what is lacking is the modal dimension, the possibility that ones claim might be ranging from claims of a priori necessary (as Gibbard wants) to a posteriori contingent. This is what I mentioned above where Gibbard made his

claim that we are committed to a priori identifying. I'd say we are not committed to it, even though we might accept it. If Gibbard is only out to characterise his line of expressivism, though, I have no trouble with his account. I only want to draw attention to the fact that is not a neutral way of doing moral semantics.

So we might want to add a fourth dimension for modal status. My suggestion for a third dimension could be included as a fourth dimension in the Gibbard network, but I doubt that it would be enlightening. For one thing, it would be extremely hard to draw (it would have to be extended in time, perhaps).

So does this mean that my claim, that what is needed is a third dimension to determine which two-dimensional network (H) should be thought of as having, fails to be neutral as well? Have I tricked my way to a three-dimensional matrix by fixing one of the dimensions? Do we need four dimensions to characterize moral claims neutrally? The expressivist that accept my argument (i.e. not Gibbard, presumably) would say so, the naturalist or realist probably would not, as they would say that the expressivist are profoundly mistaken. I was only concerned with sentences like (H) considered as true. These, the Gibbard-type expressivist would say, are cases where the plan is the ego-hedonic plan, i.e. where Gibbards third dimension is fixated. But accepting an ego-hedonic plan is not them same thing as to say that the 'rightness role' is played by the ego-hedonic. The latter is *not* a matter of settling what to do. But this seems to make it impossible for me to say what Gibbard says. Gibbard would probably want to be able to say that something could play the rightness role (where the 'rightness role' is couched entirely in non-normative terms) and still not be the thing to do. In that case, our claim that "being right is being what plays the rightness role" would not be true a priori. If this is correct, or even if it just *might* be correct, a neutral semantic characteristic should be able to represent this possibility. To do so, we might need four-dimensional equipment.

We can restrict our characteristic of moral claims to three dimensions for different intents and purposes or even two-dimensions for even more restricted intents and purposes. Or one dimension if all we want is the reference or the primary intension. This strategy is probably advisable, pedagogically and visually for all characteristics with more than two dimensions (I tried making three-dimensional truth-valuing "boxes" for moral claims, but they turned out such eye-sores that I decided to skip them here). The "diagonal" for a four-dimensional matrix would yield nothing interesting, for example.

References:

Chalmers, David "Two-dimensional" (forthcoming)

Dreier, James, "Meta-ethics and Normative Commitment"

Gibbard, Alan, "Thinking how to live"

Jackson, Frank, "From Metaphysics to Ethics"