MIT Department of Philosophy
2009 September 25

Concepts of Concepts as Normative

Allan Gibbard

Department of Philosophy
University of Michigan, Ann Arbor

I have long advocated a view of the nature of normative concepts. Equivalently, I have had a theory of the meanings of normative terms, such as ‘warranted’ or ‘a reason to’. Words voice concepts, we can say, and the meaning of a word is the concept it voices. The view I put forward is a form of expressivism. Like a non-naturalist, I don’t give straight definitions of normative terms in naturalistic terms. I do, though, claim that there is more to be said to characterize normative meanings. Normative beliefs, I said in my 2003 book *Thinking How to Live*, are states of planning, or more generally, restrictions on states of planning. That is how I explain the nature of normative concepts or the meanings of normative terms.

The concept of meaning, though, is itself puzzling. Saul Kripke famously offers an interpretation of Wittenstein, and constructs a paradox on Wittenstein’s behalf. What fact about me, he challenges us, is the fact that by the plus sign yesterday, I meant addition, and not a strange and gerrymandered arithmetic operation he calls “quaddition”? Quaddition he defines so that it coincides with addition for all numbers I take myself to have added, and even for all of the numbers I have dispositions regarding—which must, after all, be a finite set. The fact of what I meant can’t be that I was disposed to give the sum, because I was likewise disposed to give the “quum”, the result of quaddition. He then offers a diagnosis: The tie between meaning and what I do, he says, is normative: if I mean addition by the plus sign, then I should answer with the sum.

There are various ways one might take Kripke’s observation, but the one I’ll experiment with is this: that the concept MEANING is itself a normative concept.[1] Normative concepts, we might join Wilfrid Sellars in saying, are concepts “fraught with ought, or infused with the notion of reasons to do things. Prime examples are ethical concepts. The term “normative” is of course a technical one, and it has many meanings, but a central one is this: There is something puzzling about ethical concepts, as G.E. Moore emphasized. These puzzling features are shared by such concepts as that of a reason to believe something. Talk of what we should do or what we should believe doesn’t easily translate into terms suitable for empirical science. If, then, there is a single feature these concepts all share and that accounts for this puzzling nature they share, then

[1] I join Paul Horwich in using small caps to denote concepts.

whatever this feature is amounts to being normative. The concept of meaning, as I'm taking Kripke's observation, shares a crucial and puzzling feature with ethical concepts. The meaning of a term like 'plus' is somehow infused with how we ought to use the word.

On my own view, drawn from Ewing, there is a basic normative concept, involving a kind of ought or a notion of warrant. If I say that an action is admirable, I'm saying that admiring it is warranted. If I say that a claim is credible, I am saying that believing it is warranted. It is this basic notion of warrant that I explain expressivistically. In my 1990 book *Wise Choices, Apt Feelings*, I called the state of mind expressed by normative assertions "accepting a system of norms". (This was a first approximation, but I'll skip over the refinements.) In my 2003 book *Thinking How to Live*, I spoke of the state of mind as something like having a plan. Take an example: Belief in the theory of natural selection is warranted for us, in light of our evidence, but would not have been warranted for Hume in light of his evidence. When I say this, then roughly, according to my theory, I am expressing a plan to believe in natural selection for the case of having our evidence, but not for the case of having Hume's evidence.

Perhaps, then, the concept MEANING is normative, in that the concept WARRANT figures in it. Question of warrant are questions of what to do or believe or the like. So questions of a word's meaning might be questions, in some sense, of how to use the word. They aren't questions of how we do in fact use the word; that would make questions of meaning questions of empirical fact. They are questions of how *to* use the word. That is the hypothesis I want to elaborate and scrutinize.

Still, doubtless, the facts of how we do use a word will be highly relevant to what the word means. Consider a parallel: with a normative concept like good, even though, if non-naturalists and I are right, the concept isn't naturalistic, still, we can ask what natural qualities being good consists in. Likewise with the concept meaning: the concept isn't naturalistic, if I am right, but we can ask what natural features of a word and its use make a word mean what it does. Systematic answers to questions like this will comprise a theory of meaning and its bases. The theory, though, even if correct, won't be correct simply in virtue of the meaning of our word 'meaning'. What theory of meaning is correct will be a substantive question, not settled purely by what the word 'means' means, just as what theory of goodness is correct is a substantive ethical question, not settled purely by what the word 'good' means.

There is a kind of circularity in the kind of treatment I'll be examining, and a chief thing I'll be doing in this paper is to examine the nature of this circularity. Can my theory apply to itself and explain its own meaning? Or does the circularity vitiate the purported theory? If the circularity is virtuous rather than vicious, does it leave us with a genuine form of expressivism, distinct from non-naturalism for warrant and meaning?

I sketched at the start of this lecture how the things I am saying apply not only to the concept of meaning, but to the concept of mental content, of what a person is thinking. When Quine's exotic subject exclaims "Gavagai!" we can ask what he means. We can ask also what thought he is expressing. He is expressing, perhaps, the thought THAT'S A RABBIT, and he might have believed this without expressing his belief. I'll continue to speak of meaning and mental content pretty much interchangeably. I'll take the notion of expressing a state of mind as understood, and say that the meaning of an assertion is the content of the mental state that its words serve to express.

The rough idea I want to sketch, then, is this. First, the meaning of a word is a matter of how one ought to use it. Questions of what a word means are questions of how one ought to use it. Second, questions of what one ought to do are questions of what to do. One comes to a view on an ought question when one comes to a relevant plan. So questions of what a word means are questions of how to use it. What sorts of use are in play here? The meaning of a word, I'll say, is a matter of what beliefs to have as couched with that word. Here and in much else that I'll be doing, I borrow from Paul Horwich, who has a quite different, descriptivistic and naturalistic account of what the word 'means' means.

To believe a thought, I'll say more or less with Horwich, is to accept a sentence in one's own language that means that thought. For me to believe CAMBRIDGE IS BEAUTIFUL, for instance, is for me to accept my sentence 'Cambridge is beautiful', since that sentence means that Cambridge is beautiful. If Pierre speaks French and nothing else, then for him to believe CAMBRIDGE IS BEAUTIFUL is for him to accept his sentence 'Cambridge est beau,' since that sentence, in his mouth, means CAMBRIDGE IS BEAUTIFUL. I'll also follow Horwich in designating meanings by the words in my own present language that have that meaning, put in small caps. Trivially, then, my present sentence 'Cambridge is beautiful' means CAMBRIDGE IS BEAUTIFUL. I call the meaning of a full sentence a *thought*. ('Proposition' is another word that might be given this meaning, but it is often used to mean something somewhat different that pertains to reference.) Pierre and I thus believe the same thought, CAMBRIDGE IS BEAUTIFUL, when we accept our respective sentences with this meaning.

The meaning of a word in a person's language, then, is a matter of what sentences with that word the person ought to accept. And what sentences the person ought to accept is a matter of what sentences to accept if one is that person. Ought beliefs are something like plans for what to accept given the evidence and proclivities of that person. Whether to accept a sentence in one's language, though, depends on more than one what the sentence means. It depends on one's evidence. In explaining whether and why a person ought to accept a sentence, we have somehow to separate out the role of meaning. Quine was famously skeptical of any clear distinction between meaning and substance. We have to ask whether taking meaning normative addresses this problem. I'll be claiming that, to a degree, it does.

Synonymy

As Quine stresses, we will have what we need from an account of meaning if we understand synonymy, or meaning the same. A Frenchman's word 'chien' means the same as my word 'dog'. If we can explain what 'synonymous' means, we have what we need by way of saying what 'meaning' means. The French word 'chien' means DOG, in that it is synonymous with my word 'dog', and trivially, by deflation and the small caps convention, my word 'dog' means DOG. I adopt all this from Horwich.

Pierre speaks French, let's suppose. His sentence 'Les chiens aboient,' then, means DOGS BARK. What does this claim of meaning mean? I want to interpret it as a claim about how Pierre ought to use his words, about which of his sentences he ought to accept in what circumstances. The oughts that apply to Pierre and the words and syntactic devices of his sentence 'Les chiens aboient' correspond, in some crucial, meaning-determining way, to the oughts that apply to me and the words and syntactic devices of my sentence 'Dogs bark.'

I'll follow Horwich, then, in another way. The meaning of a sentence, I'll take it, is composed of the meanings of its words and syntactic devices. The structure of its meaning is its syntactic structure. What I'll want to do, however, is to take the approach of Horwich normative. Horwich's descriptive, scientific account may be the right substantive theory of what comprises meaning, but the meaning of 'meaning', on the approach I'm experimenting with, is something normative. The concept MEANING is a normative concept.

Horwich's treatment of meaning, I recognize, is not the predominant one. More prevalent treatments of meaning take the notions of truth and reference as basic. I choose Horwich's account to modify because accounts more standard among philosophers treat truth and reference as primitive. Davidson and others have a lot to say about how to attribute truth and reference, but Davidson himself insisted that the concept of truth can't be further explained. Moreover, many phenomena in thought and language don't involve reference literally, but believed reference or feigned reference, as with phlogiston and Sherlock Holmes. A theory of meaning carries the burden of somehow explaining truth and reference, but I don't want to take these as understood from the start. Horwich provides devices for thinking of meaning without supposing at the start that we understand truth and reference.

A problem for both normative and descriptive theories is how to distinguish the special explanatory role of meaning. For a descriptive theory like Horwich's, the question will be how to distinguish the role of meaning from the role of other factors in explaining linguistic phenomena like which sentences are accepted in what evidential circumstances. Quine famously maintained that this problem could not be solved. For an account of meaning as a normative concept, the problem will be to distinguish the role of meaning in explaining not why we do accept the sentences we do, but why we ought to accept certain sentences and not others.

Carnap, Russell, and Horwich have a way of isolating the meaning of a word—at least for theory-laden terms, for words that draw their meanings from their role in a theory. The crucial device has come to be called a Carnap sentence. Horwich uses the example of the term ‘phlogiston’, which got its meaning from the way it figures in a theory of burning, rust, and the like. Let *T*(phlogiston) be the theory in which the word ‘phlogiston’ figured. (Other terms as well may get their meanings from the ways they figure in the theory, but I’ll ignore them.) We ourselves don’t accept the theory, but if we study the matter, can understand what the word ‘phlogiston’ meant. What we accept is the Carnap sentence, which isolates the meaning of the term. We get the Carnap sentence as follows: First construct the Ramsey sentence for phlogiston theory, $\exists x\, T(x)$. The Ramsey sentence says, in effect, that there is something that fits what phlogiston theory attributes to phlogiston. It gives the substance of phlogiston theory, but without using the term. The meaning of the word ‘phlogiston’ is what’s left. Conditionalize phlogiston theory on its Ramsey sentence, and we have the Carnap sentence

$$\exists x\, T(x) \to T(\text{phlogiston})$$

The Carnap sentence is one we can all accept just in virtue of understanding the word ‘phlogiston’, whether or not we accept phlogiston theory. Also, we understand that if the Ramsey sentence is false, then nothing is phlogiston:

$$\neg\exists x\, T(x) \to \neg\exists y\, \text{phlogiston}(y)$$

Once we take the concept of meaning normative, the point won’t be that we do accept the Carnap sentence, or that we would accept it in such-and-such circumstances. It is that we ought to accept it. Roughly, a word means phlogiston, in the mouth of a person Jay, if Jay ought to accept this Carnap sentence. Phlogiston theory couched with the word ‘phlogiston’, after all, is entailed by the substance of the theory apart from what the word means. It thus amounts to the Ramsey sentence, along with the Carnap sentence. What the Carnap sentence adds to the substance of the theory is the meaning of ‘phlogiston’.

We’ll have to add some further requirements to this. We’ll need to add that one ought to accept the Carnap sentence *a priori*, no matter what the evidence. One ought to accept it given any assumption that doesn’t violate the meanings of the words and devices that figure in phlogiston theory apart from ‘phlogiston’.

The story of what normative characteristic is a word's meaning, then, goes something like this. Words get their meanings in sequence. I won't try saying yet how the sequence gets its start. Once initial words get their meanings, though, further words mean what they do, in a person's mouth, in view of certain oughts that obtain. The oughts take the form illustrated by the phlogiston example. They are that one ought to accept a Carnap sentence *a priori* and under all assumptions that don't violate the meanings of words earlier in the sequence.

Note that if all this works, we have also explained analyticity. Analyticity is now a normative quality. An analytic sentence is one with this normative characteristic: anyone who understands it ought to accept it, solely in virtue of the meanings of the terms involved. An analytic contradiction is a sentence such that anyone ought to reject it solely in virtue of what it means. Meanings are given by a sequence of Carnap sentences, which one ought to accept *a priori* and given any assumption not logically contradicted by the Carnap sentences earlier in the sequence.

All this, as I said, leaves the question of how the sequence gets started. It starts with syntactic devices such as predication, and with basic logical words like 'and' and 'not'. It also starts with basic empirical terms, terms whose meanings are tied directly to experience. I don't have anything like a full story of how this goes, but examples are suggestive. Crucial, for instance, to the meaning of a conditional is that *modus ponens* applies. One ought not to accept the conditional and its antecedent and reject the consequent. This applies no matter what else one ought to accept or reject. A term means IF . . . THEN only if it has this normative quality.

As for basic empirical terms, the word 'dog' gets its meaning in part from its tie to paradigm dog experiences. If one's words mean what they do in English, then one ought to accept one's sentence 'That's a dog' when one has an experience as of a dog in plain sight right in front of one. This ought isn't indefeasible, since one might have reason to expect a sheep in dog's clothing. But some ought like this, I suspect, helps characterize what the word 'dog' means.

I don't claim to have these parts of the project carried through properly. It does suggest a strategy, though, for how we might understand the concept of meaning as a normative concept. It isolates the role of meaning in explaining why we ought to accept a sentence. It also lets us explain the concept of analyticity as a normative concept.

Why might taking the concept of meaning normative help us solve problems whose solutions are elusive when we take the concept to be a naturalistic one? One problem that plagues naturalistic concepts of meaning is the problem of error. People make mistakes, and so what a person is disposed to do needn't be in accord with what his words mean. He may bollix his arithmetic, for instance, and still mean addition by the plus sign. The answer he ought to accept, though, will tie in with its meaning.

This still allows that he means what he does in virtue of some feature of the pattern of his dispositions to accept or reject sentences with the word. The story now, however, goes like this: In virtue of his dispositions to accept and reject sentences, he ought to accept certain sentences in certain conditions, and to reject others. Oughts, after all, are based on is's: one ought to take an umbrella if it's raining and one would be wet and miserable without it. Something in this pattern of oughts for accepting sentences is what it is for a word to have its meaning. If Pierre's word 'chien' means DOG, that may well be because of his dispositions to accept or reject sentences with the word 'chien' in them. The claim that his word 'chien' means DOG then amounts to this: the sequence of oughts that gives the meaning of his word 'chien' matches the sequence of oughts that gives the meaning of my word 'dog'.

Beliefs about Meaning as Plans

Ought beliefs, I keep saying, are plans, and epistemic ought beliefs—beliefs as to what one ought to believe—amount to plans for belief. If MEANS is a normative concept, "fraught with ought", the beliefs about meanings likewise amount to plans, along with restrictions on plans, restrictions on restrictions, and the like.[2] Beliefs about meanings involve contingency plans for accepting sentences in one's own language and in other languages. How will this work? In particular, what is involved in interpreting the meanings of sentences in a language not one's own? In what follows, I assume, vaguely at first, that ought beliefs are plans, and explore this question.

Pierre, again imagine, speaks only French, whereas I speak only English. I believe that dogs bark, and for the case of being Pierre, with Pierre's evidence, I plan likewise to believe that dogs bark. This, we are supposing, is what it is for me to believe that Pierre ought to believe that dogs bark. What, though, is the nature of this plan? It can't, it seems, be a plan to accept my sentence 'Dogs bark,' since Pierre doesn't have that sentence available to him—and if he did accept it, still, in his mouth, it might not mean DOGS BARK. We might try saying that my plan, for the case of being Pierre, is to believe DOGS BARK, which is a thought and not a sentence. Saying this wouldn't work, however: We explained believing a thought, after all, as accepting a sentence in one's language that has that meaning. For my own case, for instance, believing DOGS BARK consists in accepting my sentence 'Dogs bark,' which, trivially, means DOGS BARK. For Pierre's case, we will want to say, it consists in accepting his sentence 'Les chiens aboient,' which likewise means DOGS BARK. Not yet, though, are we in a position to say this, since the question we face is what it means to say that Pierre's sentence 'Les chiens aboient' means DOGS BARK.

[2] See my *Thinking How to Live* (2003).

Here is a possible answer, which I will develop: I plan for the case of being Pierre forthwith.[3] This is a case of being myself with my own properties, but expecting momentarily to be Pierre with his. My question is what to believe for the circumstance I am about to be in, and I couch my answer in my own language. For this hypothetical case of being Pierre forthwith, I plan, for instance, to believe of myself as I am about to be, that I speak only French. My plan is to believe this not of myself as I am, untransformed, but of myself as I am about to be.

What is it, then, to plan a belief for a circumstance? For a speaker of English like me, what is it to plan, say, to believe DOGS BARK for the case of being Pierre? It is to plan to accept one's sentence 'Dogs bark' for myself as I will forthwith be, in the hypothetical case where I will forthwith be Pierre with all his properties.

To elucidate the concept of meaning, we will need another kind of plan as well. We'll need a sense in which, for the case of being Pierre, I plan to accept the French sentence 'Les chiens aboient' and not to think in English. Fortunately, this kind of plan is more straightforward. My plan is this: once I am Pierre, to accept what will then be my sentence, 'Les chiens aboient.'

On what features of Pierre might I base this plan? I might base it on Pierre's linguistic proclivities, on his dispositions to accept or reject sentences of his in various evidential circumstances and conditional on various other plans. I might base it, in other words, on the sorts of dispositions that Quine takes as grounds for a radical translation. One way I might try putting this is: "If I'm Pierre and I know that my dispositions are such-and-such, let me accept my then sentence 'Les chiens aboient.'" (This needs more explanation, since I haven't yet glossed talk of what I "know", but leave that aside for now.) Which such plans to have is a question not for the metatheory of meaning that I am constructing, but for a substantive theory of meaning and its bases. In a strict sense, then, it is beyond the scope of this paper. As an example, though, I can indicate the sorts of plans that I myself have, and hence the kind of substantive theory of meaning that I find plausible. This substantive theory of meaning will be much like Horwich's. I plan to accept my sentence 'Dogs bark' for my own evidential circumstances and for those of almost everyone who is familiar with dogs. These plans extend to any circumstance where I have certain sorts of dispositions: to accept 'That's a dog' with my attention focused on a dog in plain sight, to accept 'It's barking' when I focus my attention on an animal I observe clearly barking, and the like. My plans for the case of being Pierre are similar, but with 'chien' substituted for 'dog', with 'aboyer' in place of 'bark', and the like. Indeed for the case of being a person whose dispositions to accept sentences fall in a pattern like this, as my own dispositions do, my plans correspond to my plans for my own case as an English-speaker.

[3] I draw this idea of hypothetically being forthwith in a circumstance to R.M. Hare, who puts his account of moral convictions in terms of one's preferences for the case of being forthwith in someone else's shoes. [See *Moral Thinking* where? xxx.]

In light of certain ways that Pierre's dispositions correspond with my own, though with 'chien' in place of 'dog' and the like, I plan to accept certain sorts of sentences in a corresponding way—again with 'chien' in place of 'dog' and the like.

My plans, then, are, as I said, of two kinds. Pierre, imagine, sees a dog right in front of him, in epistemic circumstances that are in no way unusual. 1) For the case of being about forthwith to be Pierre, I plan to accept my sentence 'That's a dog' regarding the circumstance I'll then forthwith be in. 2) For the case of already being Pierre, with Pierre's dispositions to accept and reject French sentences, I plan to accept the sentence, which will then be mine, 'Voici un chien.' Part of what is involved in believing that his word 'chien' means DOG is for these two kinds of plans to match in this way, with 'dog' in one matching 'chien' in the other.

Using these materials, we can construct complex plans. Judgments of what things mean, the hope is, will consist in complex patterns of planning. My broad question is whether planning has the resources to capture ought beliefs, and in particular, beliefs about meanings taken as complex ought beliefs. The planning states I describe in what follows, I regret to say, get almost intractably complex, and it will be good to get back to the more compact language of 'ought' and 'means'. I want to know, though, whether such language can be interpreted expressivistically, in terms of states of planning and restrictions on states of mind.

An example: MASS

Quine famously rejected the analytic-synthetic distinction. We can ask whether such a distinction becomes intelligible once we take meaning normative. I'll use a somewhat arcane example, but one where the structure of the problem is especially clear: my word 'mass' as used in Einstein's special theory of relativity. 'Mass' is a theoretical term, in that it gets its meaning from a theory in which it figures. The problem is this: In traditional Newtonian physics, the mass of a body is taken to be constant, and not to depend on its velocity. So long as this is so, two claims of Newtonian theory come out equivalent.

Force equals mass times acceleration.

Force equals the rate of change of momentum.

These are equivalent, because momentum is defined as mass times velocity, and so if mass is constant, the rate of change of momentum is just the mass times the rate of change of velocity—which is the acceleration. In Einstein's theory, mass increases with velocity, and so these two are no longer equivalent. In Newtonian theory, we can use either of these formulas to characterize mass, but in Einstein's theory they characterize different magnitudes. Let's suppose we believe Einstein's theory. We can no ask what Newtonian physicists meant when they used the term 'mass'. Some say it's the one, and some say it's the other. Thomas Kuhn thought they meant something incommensurable with our terms. Quine thought there's no fact of the matter. I won't try to say what answer is right, but I will sketch an account of what's at issue among the different views.

I'll imagine the particular way our meanings work in Einstein's theory as follows: the words 'force' and 'velocity' get their meanings first. Next the word 'momentum' gets its meaning from the theory, 'Force equals the rate of change of momentum.' Finally, we reach the step I will scrutinize: let the theory *T*(mass) that gives the word 'mass' its meaning be

> Momentum equals mass times velocity.

Apart from the word 'mass' itself, then, this theory *T*(mass) is couched in words in my language whose meanings are already established and not in question. Given these meanings, I said, the meaning-determining pattern of oughts that gives the meaning of my new word 'mass' is that one ought to accept the extended Carnap pattern:

$$\exists x\, T(x) \rightarrow T(\text{mass})$$
$$\neg\exists x\, T(x) \rightarrow \neg\exists y\, \text{mass}(y)$$

The pattern, to determine my meanings, must obtain under any meaning-compatible supposition couched in words already given meanings. It must also obtain given any plan that doesn't go against the meanings of these terms or involve the word 'mass'. I'll fix on a particular case of this invariance in the form of three suppositions. The first is a normative assumption that is clearly of no relevance, and I won't mention it further in the physics example. a) One ought to pursue perfection. The next is evidential. b) The total evidence is *E*, where *E* is the total evidence available to physicist Hans in 1880. The third is in language already given meaning. c) For a given object, momentum is proportional to velocity. Thus in short, a) is a planning supposition, b) an evidential supposition, and c) a supposition in language already given meaning. Our aim is to understand what sort of planning constitutes thinking that one ought to accept something under each of these three kinds of suppositions.

Start with plans to accept sentences in one's own language under suppositions of these kinds. Hypothetically planning to accept $\exists x\, T(x)$, I plan to accept *T*(mass). This planning pattern, moreover, is invariant under suppositions a)–c).

Now let's turn to the case of Pierre, a Newtonian physicist in 1880 who spoke and thought in French. He has words 'force', 'moment', and 'masse', and I particularly wonder whether by his word 'masse' Pierre meant MASS. What does this issue consist in? What would it be for me to think that by his word 'masse', Pierre meant mass? It consists in believing, among other things, that this same ought pattern—one that contributes to a word's meaning MASS—obtains in the case of Pierre. In particular, it depends on what sentences Pierre ought to accept given the evidence that supports Einstein's theory and the Ramsey sentence for Einstein's special relativity. If by 'moment' he means MOMENTUM and by 'masse' he means MASS, then for this epistemic circumstance, he ought to accept his sentences,

> La force egale le taux de changement du moment;
>
> Le moment egale la masse foix la vitesse.

If he means something else by it, such as the rest mass which doesn't change with the velocity, then under these same assumptions, he should reject these sentences and accept,

> La masse est constante quand la vitesse change.

What Pierre ought to accept by way of sentences in his own language depends on certain facts about him, such as his proclivities to accept or reject sentences in various circumstances. These proclivities are matters of his somewhat messy dispositions to accept or reject sentences. Are the facts of what these proclivities are available to him? He can engage in thought experiments, accepting and rejecting sentences with regard to circumstances that he can imagine, and from such thought experiments he can glean much information. Much about one's linguistic proclivities is thus available to a person on the basis of such thought experiments. I don't regard it as clear, however, that all aspects of one's linguistic proclivities that bear on what one means are subjectively available to one. In a sense, then, which of Pierre's French sentences he ought to accept for an evidential situation may not be a matter of the pure subjective oughts of the matter—where by "subjective oughts", I mean oughts in light of the person's own evidence, not in light of truths he has no way of knowing. Which sentences he ought to accept depends in part on facts about his linguistic proclivities that he might not have available to him. We must see if we can fashion plans that incorporate all this.

Can we explain this sort of ought by saying what sorts of plans constitute believing in such an ought? Matters rapidly become complex, but here is an attempt: Let the question be whether Pierre, in light of his linguistic proclivities *L*, ought to accept a sentence *S* of his for the case of his total evidence being *E*. (*E* need not be evidence that Pierre already has.) To believe that he ought to, I'll try saying, is to plan as follows: One plans for the case of forthwith being Pierre with all his properties, hypothetically planning to believe of oneself everything true of Pierre—including the facts of his linguistic proclivities. In this frame of mind, one settles whether, as Pierre, to accept sentence *S* of Pierre's language for the case of one's total evidence being *E*.
I might, for instance, ask about Pierre's sentence 'Voici un chien,' for the typical evidential circumstance of seeming to see a dog in plain sight right in front of one. I then ask myself this planning question: Suppose I am about to be Pierre, with all his linguistic proclivities, and I can

settle whether, as Pierre, to accept one's sentence 'Voici un chien' for evidence *E*. Shall I settle on accepting, as Pierre and for evidence *E*, the sentence 'Voici un chien'?

In so planning, I am in my actual state, a speaker of English. I contemplate:

> The hypothetical state of rightly expecting forthwith to Pierre, with all his characteristics. (When I speak of expecting "rightly", this is a planning restriction: I condition on a plan to believe, in this state, everything that is the case with Pierre.) In this state, my language is English.
>
> The hypothetical state of already being Pierre (and so speaking only French).
>
> Evidence *E* as one's total evidence.

I consider states of Pierre that consist in accepting a French sentence *S* for a totality of evidence *E*. I plan whether, as me expecting to be Pierre, to settle on accepting, as Pierre, sentence *S* for evidence *E*. (Here and from now on, I ignore the further conditions I included in my list a)–d), in order to highlight aspects of what I am saying that are already complex.)

The question for Pierre himself is just what to believe given evidential state *E*. He might put one of his beliefs to himself, 'Que la force egale la masse fois la vitesse.' The question for me expecting to be Pierre is what to accept once I am Pierre. I conceive of my alternatives as sentences such as 'La force égale la masse fois la vitesse.' My question is which such sentences to accept, once I'm Pierre, for totality of evidence *E*. The question for me as I am is what, in a hypothetical frame of mind, to plan for this case of expecting forthwith to be Pierre, given a plan to believe, then, everything that is true of Pierre.

If all this is right, it gives us a normative distinction between what's analytic and what's synthetic, both in my language and in Pierre's. What's analytic is what one ought to accept on any supposition that doesn't go against the meanings of one's words. The meanings of one's words are given by a sequence of extended Carnap sentences that one ought to accept given any supposition that doesn't go against the meanings of words prior in the sequence. In that sense, they are sentences that one ought to accept come what may. I'll explain later how this is a stronger requirement than being *a priori,* being such that one ought to accept it for any conceivable evidential circumstance. Quine thought there might be no fact of the matter whether Pierre means MASS by his word 'masse'. Pierre, after all, on contemplating the evidence for Einstein's theory and its Ramsey sentence, might just be confused about what to say and think for that far-fetched case. I say there is a quasi-fact. What Pierre means is a matter of which of his sentences he ideally *ought* to accept for such cases as this one. Obviously, what he ought to accept depends on such natural facts about him as his linguistic proclivities. People who agree on these facts, though, can disagree on their import for what he means. In that case, the disagreement is over what sentences in his language, in light of his linguistic proclivities and the like, Pierre ought accept given certain assumptions. I have tried to illustrate how this might work in a particular example.

Normative Concepts

I have been sketching how we might try combining two claims:

> 1) The concept of meaning is a normative concept.
>
> 2) Normative concepts are to be explained expressivistically, by saying what states of planning constitute believing a normative claim.

I asked first how we might understand claims about what a person's words mean as normative claims, which we can couch with a primitive concept OUGHT. This included how to interpret claims about analyticity, in a person's language, in normative terms. This is equivalent to explaining conceptual relations among thoughts, such as inconsistency and entailment. I then sketched how ought claims might be explained obliquely, in terms of the planning states that accepting them consists in. Finally, I took an example in the history of science for which claims of analyticity have seemed problematic. I constructed a test case: a complex normative claim that must be true if a French-speaking Newtonian physicist meant MASS by 'masse'. I asked about the sorts of plans that constitute believing this normative claim.

How are we to understand my theory itself, my claim that 'ought' means what I say it does. The claim amounts, I say, to a plan for how to use the word 'ought', in light of our linguistic proclivities. I'll have persuaded you if I convince you to accept a certain plan for believing or disbelieving the thoughts we couch with the word 'ought'. The plan is to do the following two things equivalently: (i) accept or reject my sentence 'I ought not to kick that dog' and (ii) adopt or spurn the plan LET ME NOT KICK THAT DOG. It is to do this given any plan, assumption, and evidence compatible with meanings that are already explained. To see if you believe my account of normative meanings, see if your plans for belief and sentence acceptance take that form.

This framework explains Moore's refutation of simple forms of analytic naturalism for ethical terms. A hedonistic egoist, suppose, says that 'ought' just means WOULD BE MOST GRATIFYING. To conclude whether he really means this by his word 'ought', I think what sentences of his to accept for cases like the following: expecting forthwith to be he and planning not to torment chickens even when doing so would be most gratifying. For that case, I myself might plan to settle on rejecting his sentence, 'One ought to torment the chicken." If I do so plan, I reject his claim as to what he means by 'ought'. On the other hand, I plan to accept his sentence given the hypothetical plan to always to do whatever is most gratifying. All this fits my account of what 'ought' means, and amounts to a kind of argument that Moore gave against such naturalistic definitions of ethical terms.

All this is quick, sketchy, and perhaps overly complex. If it works, though, it might show us how the concept of meaning isn't naturalistic. Is it nonnaturalistic, though, in the way that G.E. Moore thought that the concept GOOD is nonnaturalistic? I argued in my 2003 book *Thinking How to Live* that Moore's arguments against analytical naturalism in metaethics support not the view that good is a non-natural property, but that the concept GOOD is non-naturalistic—that it is indefinable in naturalistic terms. Expressivists accept this as far as it goes, but protest that it leaves too much unexplained. What could a non-naturalistic concept be, and how could such a concept be legitimate? There's a further way to explain the concept WARRANT, I say as an expressivist: by its tie to motivation.

One thing I have said about this tie, though, strikes many as too strong: that if a person believes she ought to do a thing, that any alternative is unwarranted, then she does it; if she decides otherwise, she doesn't really and fully believe that she ought to do it. Seeming acrasia is a rapid change of mind as to what one ought to do. I now want not to insist on this, but to press two other points connected to the expressivist tradition which starts with Barnes and Ayer.

First, we can explain the concept WARRANT in a way that is naturalistic, in a way, but obliquely. Instead of giving a naturalistic synonym, we can specify in naturalistic terms what believing an act warranted consists in. What it consists in, moreover, will among other things be a strong tendency to do as one thinks warranted. Expressivists, though, have been vexed to explain the status of the claim that meaning WARRANTED by a word consists in meeting such-and-such a naturalistic specification. I now say that this claim for a naturalistic theory of the concept WARRANT is itself normative, that the concept WARRANT figures in it.

Second, there is a stronger tie of the concept WARRANT to action, a tie to action and the like that distinguishes normative concepts from naturalistic concepts. The tie is conceptual. Just as it is contradictory, conceptually, to believe the premises of a valid argument and disbelieve the conclusion, so it is contradictory to believe an act unwarranted but to decide on it anyway. In this sense, unwarranted implies Don't! Conceptual contradiction we explain in the way I have been sketching; the claim that a person is in a contradictory state of mind is itself a normative claim. The claim is that on the widest intelligible range of suppositions, one ought not to believe that one ought to do a thing and decide otherwise.

All this is very quick and unargued for, but I'll leave it at that, and just make a final observation. This is a position that a non-naturalist could accept. Indeed, I read T.M. Scanlon as accepting it. If this is right, then the naturalistic and expressivistic traditions in metaethics can converge. Not all non-naturalists and expressivists with accept the resulting account of normative and semantic concepts, but there is a way for both camps to arrive at equivalent positions.

References

The following are some of the principal sources and background:

Brandom, Robert (1994). *Making It Explicit* (Cambridge, MA: Harvard University Press).

Gibbard, Allan (1990). *Wise Choices, Apt Feelings: A Theory of Normative Judgment* (Cambridge, Mass.: Harvard University Press).

Gibbard, Allan (1994). "Meaning and Normativity". *Philosophical Issues* **5**, Enrique Villanueva, ed., *Truth and Rationality* (Atascadereo, CA: Ridgeview Publishing Co., 1994).

Gibbard, Allan (1996). "Thought, Norms, and Discursive Practice: Commentary on Robert Brandom, *Making it Explicit*". *Philosophy and Phenomenological Research* **56**, 699–717.

Gibbard, Allan (2005). "Truth and Correct Belief". *Philosophical Issues* **15**: *Normativity*., 338–350

Gibbard, Allan (2008). "Horwich on Meaning" (2008). *Mind* **117**:465 (January), 141–166.

Horwich, Paul (1998). *Meaning* (Oxford: Clarendon Press).

Horwich, Paul (2005). *Reflections on Meaning* (Oxford: Oxford University Press).

Kripke, Saul (1982), *Wittgenstein on Rules and Private Language* (Cambridge, Ma.: Harvard University Press).

Scanlon, T.M. (2007). "Structural Irrationality". G. Brennan, R. Goodin and M. Smith (eds.), *Common Minds: Themes from the Philosophy of Philip Pettit* (Oxford: Oxford University Press), pp. 84–103.