Does Semantics Need Normativity?

Comments on Allan Gibbard, *Meaning and Normativity*

Åsa Wikforss
Stockholm University

Forthcoming in *Inquiry*

In the book Gibbard proposes, first, that statements about meaning are normative statements and, second, that they can be given an expressivist treatment, along the lines of Gibbard’s preferred metaethics. In my paper I examine the first step: The claim that meaning statements are to be construed as being normative, as involving ‘oughts’. Gibbard distinguishes two versions of the normativity of meaning thesis – a weak version, according to which every *means* implies an *ought*, and a strong version, according to which for every *means*, there is an *ought* that implies it. I argue that neither thesis withstands scrutiny. The weak thesis depends on assumptions about the notion of semantic correctness that the anti-normativist rejects, and the strong thesis does not solve the problems Gibbard wants it to solve: the problems of indeterminacy and meaning skepticism. I conclude that semantics does not need normativity.

The role of norms in thought and action has emerged as a central theme in contemporary philosophy – not only in the areas where norms have traditionally taken center stage, such as ethics and aesthetics, but also in the philosophy of mind and language. Linguistic meaning, mental content, belief and reasons have all been claimed to essentially involve norms. It is only natural, then, that Allan Gibbard, who has explored in depth the nature of normative discourse, has now turned his attention to linguistic meaning and mental content. In his richly argued and intriguing book he spells out how the expressivist theory that he has applied to ethics can also be applied to semantics, and the implications that holds for central issues in philosophy of language, such as analyticity, meaning skepticism, naturalism about meaning and indeterminacy. As Gibbard exclaims: “If meaning is normative, then a central topic in the philosophy of language becomes a part of metaethics. Metaethics can turn imperialistic, and grab territory from the philosophy of language” (pp. 7-8).¹

This is a conditional claim and extending normativism beyond the obvious domains is not uncontroversial. On the contrary, a number of people have provided arguments

¹
against such an extension. Meaning normativism has been questioned (Bilgrami 1993, Boghossian 2005, Glüer 1999, Wikforss 2001, Hattiangadi 2006 & 2007, Glüer & Wikforss 2009a), as well as content normativism (Glüer & Wikforss 2009a), and normativism about belief and epistemic reasons (Bykvist & Hattiangadi 2007, Glüer & Wikforss 2013). Gibbard is not unaware of this, noting that some people have rejected the normativity of meaning and content as an illusion (p. 18). However, he hopes to devise a form in which the normativity thesis might be defensible. In my comments I shall focus on whether he succeeds in this: How does meaning normativism, as defended by Gibbard, fare with respect to the anti-normativist arguments? I will suggest that nothing Gibbard says provides a reason for the anti-normativist to be converted, and that the extension of normativism to semantics remains unwarranted.

This means that I will not really get to the second part of Gibbard’s book, where he proceeds to spell out his own preferred, expressivist treatment of semantic statements. If statements about meaning and mental content are not normative, if they do not involve any oughts, then the expressivist analysis will not apply to them.

I begin with some preliminary remarks concerning semantic normativity, setting the stage for a discussion of Gibbard’s version of meaning normativism.

1. Semantic normativity

Since the publication of Kripke’s book on rule following a variety of theses have been collected under the slogan ‘meaning is normative’ (Kripke 1982). In our Stanford entry on the normativity of meaning and content, Kathrin Glüer and I tried to sort this out.ii To say that meaning is normative, we suggested, is to make a claim about the nature of meaning, about what meaning is: It is to say that meaning is necessarily normative, where this is to be understood as at least involving metaphysical necessity, possibly also conceptual necessity.iii
According to meaning normativism, thus, the following is necessary for an expression $e$ having a certain meaning $M$ (for a speaker $S$ at a time $t$):

$$ (M) \text{ } e \text{ means } M \text{ for } S \text{ at } t \text{ only if a norm for } e \text{ is in force for } S \text{ at } t $$

We also classified various types of norms, in order to pin down the relevant meaning norms. First, the meaning norms are standardly taken to be rules of actions (rather than ‘rules of being’, merely axiological). Second, the rules cannot be instrumental, contingent on $S$’s aims, since they are supposed to be essential to meaning. Moreover, they are typically understood as being prescriptive, as involving genuine ‘oughts’, telling us what to do. And finally, their provenance must be semantic – their force must derive from the possibility of meaning itself.$^iv$

The claim that meaning is essentially normative, we suggested further, can be understood in two ways, depending on the order of metaphysical priority: As the claim that meaning is such that it immediately implies certain oughts (remaining neutral on how meaning is determined), or as the claim that the norms metaphysically determine, or constitute, meaning. The first type of normativity thesis we called ‘ME-normativism’, for meaning engendered normativity. The idea, here, is that normative consequences can be derived from the fact that an expression (as used by $S$ at $t$) has a certain meaning $M$. This is the type of thesis, arguably, that Kripke is defending when he speaks about the normativity of meaning. The second type of normativism belongs to metasemantics or foundational semantics, since it tells us something about the determination of meaning – hence MD-normativism.$^v$ The basic idea is that meaning facts hold in virtue of certain rules being in force for the speaker. What does it mean for the rules to be in force for the speaker? The most common view is that the rules are in force for the speaker if she follows them, if she is guided by them. This version of the MD-normativity thesis we have labeled guidance normativism.$^vi$
With these distinctions in place, let us turn to Gibbard’s normativism as applied to meaning and mental content. What type of normativism does he defend and on what grounds?

2. Gibbard’s normativism

Although Gibbard is as much concerned with mental content as he is with meaning, his starting point is meaning normativism. However, there is a twist: He does not actually hold that meaning, the property, is normative. Instead, his normativity thesis is explicitly said to be only about the concept of meaning, the meaning of ‘meaning’. Gibbard therefore denies the implication from conceptual necessity to metaphysical necessity: While the concept of meaning is such that it is a conceptual truth that meaning involves ‘oughts’, it does not follow that meaning is normative. His slogan, he says, is not ‘meaning is normative’, but “the concept of meaning is normative, whereas the property of meaning is natural” (p. 25).

In a sense, then, Gibbard does not accept (M) at all (and hence does not accept either ME-normativism or MD-normativism) even though he does accept that statements about meaning have immediate normative implications. This is a position that, as far as I know, is unoccupied in the debate over the normativity of meaning. There are those (like Gibbard) who take meaning statements to have normative consequences as a matter of conceptual necessity, but they also take this to imply metaphysical necessity – i.e. they take it to imply that meaning, the property, is normative. Gibbard’s combination of normativism and non-normativism is therefore interestingly different. However, it also causes some complications, as we shall see.

Gibbard distinguishes between a weak and a strong version of meaning normativism (p. 12). According to the weak thesis every means implies an ought: From meaning ascriptions follow straight ascriptions of ought as a matter of conceptual necessity.
The strong thesis adds that the meaning of a term can be characterized in terms of patterns of oughts that apply to it, so that for every means, there is an ought that implies it. Let us start by examining the weak thesis.

2.1 The weak thesis

Gibbard’s weak thesis can plausibly be seen as a version of ME-normativism applied to the concept of meaning: The idea is that every statement about meaning very directly implies certain ought statements (even if, again, meaning itself is not normative). Gibbard also suggests that the weak thesis is the thesis defended by Kripke’s Wittgenstein. For instance, discussing Kripke’s claim that the relation between meaning addition and future action is normative, Gibbard writes: “Perhaps, then, Kripke’s thesis is that if I now mean addition by plus, I now should answer ‘125’. And that’s how I am interpreting the thesis: as the claim that means implies should” (p. 54). Moreover, this is a matter of analytic entailment – the ought follows from the meaning of ‘means’, from the concept expressed. The hypothesis that the concept of meaning is truly normative, Gibbard writes, would explain something that needs to be explained: “It would explain why certain oughts follow from meaning claims invariably” (p. 13).

Why does Gibbard think that meaning statements entail certain ought statements? One idea, it seems, is that there is an essential tie between the concept of meaning and the concept of semantic correctness conditions and, moreover, that the latter concept is normative: If I mean PLUS by the plus sign, then answering ‘125’ to ‘58+67’ is the correct answer – hence, I ought to answer ‘125’. Gibbard is quite explicit that he takes correctness to be a normative notion. Discussing the idea that for belief, correctness is truth, he writes: “Correctness, now seems normative. That is to say, the concept CORRECT seems to be a normative concept” (p. 75). And he goes on to apply the same lesson to meaning. Since
believing a thought amounts to accepting a sentence that means it, Gibbard writes, one ought to accept a sentence if and only if what it means is true: “The meaning of a sentence, we can thus try saying, is a matter of the conditions under which one ought, objectively, to accept it. Meaning thus is characterized in normative terms” (p. 76). Ultimately, according to Gibbard, this objective ought is to be defined in terms of the more primitive, subjective ought – the ‘oughts’ that are directly motivating. Ultimately, therefore, semantic correctness conditions are tied to evidence, rather than to truth. However, the basic idea is the same: From the fact that meaningful expressions must have correctness conditions (be they truth conditions or evidential conditions), it follows that the concept of meaning is normative.

This suggests that Gibbard relies on what Kathrin Glüer and I have called the ‘simple argument’ in support of ME-normativism. The argument is simple in the sense that it is pre-theoretical, independent of any particular theory of meaning, just moving from the observation that if an expression is meaningful then it has certain correctness conditions, (C), to the claim that if an expression is meaningful then this has normative consequences (N):

(C) For any speaker S, and any time t: if ‘dog’ means DOG for S at t, then it is correct to apply ‘dog’ to an object iff x is a dog.

(N) For any speaker S, and any time t: if ‘dog’ means DOG for S at t, then S ought to apply ‘dog’ to an object x iff x is a dog.

How compelling is the simple argument? It is clear that the need for correctness conditions cannot be questioned: It is indeed essential that meaningful expressions have semantic correctness conditions, whether the basic notion of semantic correctness is that of truth (as in C) or warranted assertibility, etc. Having semantic correctness conditions is simply that which distinguishes a meaningful expression, such as ‘dog’ from a mere scribble. In fact, this may even be a conceptual truth, which is what Gibbard needs. What is less clear is that the notion of semantic correctness conditions is normative. Indeed, this is the point at
which anti-normativists have attacked the simple argument. If ‘dog’ means DOG then ‘dog’ is
ture of a certain class of animals, false of others, but nothing seems to follow (directly) about
how the speaker ought to use the term.\textsuperscript{viii} We have made this point in terms of categorization:
The appeal to semantic correctness conditions merely gives us a way of categorizing
applications of ‘dog’ into two basic kinds (the true and the false, for instance), and this does
not in itself entail that the term ought to be applied in any particular way. Of course, like all
categorizations one can employ semantic categorization to generate normative consequences
– if one adds a suitable norm. It may be that we ought to speak the truth, and this norm taken


together with the fact that ‘dog’ is true of dogs only implies that S ought to apply ‘dog’ only
to dogs. But, obviously, this fails to show that meaning facts by themselves have normative
consequences (or, to put it Gibbard’s way, that statements about meanings analytically entail
ought-statements).

No doubt, Gibbard would deny all this, and insist that ‘oughts’ follow from
meaning statements as a matter of conceptual necessity. If so, he has to say one of two things:
He has to say, either, that the normativist and the anti-normativist have different concepts of
meaning, or that the anti-normativist is conceptually confused. Both options, it should be
clear, are problematic. Granting that the anti-normativist simply has a different concept of
meaning would render the normativity thesis rather uninteresting: It would just be a thesis
about \textit{one} concept of meaning, one that some but not all philosophers share. Indeed, since
Gibbard wishes to keep distinct the concept of meaning from the property picked out, \textbf{means
such-and-such}, and he grants that the normativity thesis only holds for the concept, not for
the property, the resulting thesis would seem to be very weak indeed. If meaning (the
property) is not normative, and if there is a concept of meaning (and semantic correctness)
that is not normative, then why does it matter to the project of semantics that there is another
concept of meaning that \textit{is} normative? The second option is even less attractive, since it
implies that the anti-normativist is confused about the concept of meaning. Wherein lies the confusion? After all, the anti-normativist grants that correctness conditions are essential to meaning, and she argues that there is a non-normative construal of these conditions that does what is required for the purposes of semantics. She may be wrong, of course, but it would hardly seem to be a conceptual truth that she is wrong.ix

In sum, Gibbard’s defense of the weak normativity thesis does not go very far. The simple argument is not compelling and the suggestion that normativism is needed to explain why “certain oughts follow from meaning claims invariably” (p. 13), begs the question since what the anti-normativist denies is precisely that oughts follow from meaning claims. Next, let us consider Gibbard’s discussion of the strong normativity thesis, where Gibbard does provide some more substantive considerations in support of meaning normativism.

2.2 Gibbard’s strong thesis

Although Gibbard’s stronger thesis is also put forth as a claim about the meaning of ‘meaning’, it does not simply rest on the intuition that meaning statements entail oughts. Rather, it is said to draw support from the claim that extant naturalistic theories of meaning all fail – Quine, Fodor, Millikan, Horwich have all tried to come up with naturalistic theories of meaning but have failed (pp. 16-17). To illustrate this, Gibbard spells out in some detail why he thinks that Horwich’s theory fails to eliminate unacceptable indeterminacies. The best explanation for this failure, Gibbard suggests, is precisely that the concept of meaning is normative.

According to the strong thesis, again, we should “characterize meaning in fully normative terms” (p. 12). The proposal is not just that the concept of meaning can be analyzed in terms of oughts, it is also suggested that claims about the meanings of specific expressions
can be analyzed in terms of a specific set of oughts, characterizing that particular meaning. So for every *means* there is an *ought* that implies it, since every *means* is to be analyzed in terms of certain *oughts*. For instance, Gibbard writes, the phrase ‘means DOG’ in my mouth means “being governed by the pattern of oughts that characterizes the actual meaning now of my own term ‘dog’” (p. 114). The result, it seems, is a form of normative inferential role semantics. Thus, in chapter 6 Gibbard goes on to explain the concept *meaning* in terms of its ‘normative role’, the inferential ‘oughts’ that are entailed by meaning statements. For example, he suggests, the statement “Ursula means SOMETHING by ‘something’” entails, analytically, that if she accepts ‘Snow is white’ she ought to accept ‘Something is white’.

This is a tall order. It is not just that inferential role semantics is problematic. It is also so very implausible to suggest that such a semantics falls out of the meaning of ‘means’, as a matter of mere conceptual analysis. This brings back the problem above, but with a vengeance. Are we really to believe that everyone who does not accept an inferential role semantics has a different concept of meaning (or is conceptually confused)? What is worse, even if it is granted that Gibbard’s normative version of inferential semantics follows from the concept of meaning, it is very difficult to see how it could be understood merely as a thesis about the *concept* of meaning, and not about the property. After all, semantic theories, such as inferential role semantics, tell us something about the *meaning* of our terms. Indeed, this also is what Gibbard seems to be doing when he tells us that the actual meaning of my term ‘dog’ can be characterized by a certain pattern of semantic oughts. How can this not be a claim about the *property* of meaning DOG by ‘dog’?

The trouble derives from Gibbard’s starting point, the assumption that philosophers of language are primarily preoccupied with providing an analysis of the meaning of ‘meaning’. Of course, there have been philosophers who have tried to analyze the concept of meaning, most famously perhaps Grice in his analysis of the notion of non-natural
meaning. But for the most part this does not at all seem to be what is going on in contemporary semantics, and certainly not in the writings of the people Gibbard mentions, such as Quine, Fodor, Millikan, and Horwich. The problem is accentuated when we turn to the issue of indeterminacy. Gibbard’s central argument in support of strong normativism, again, turns on the idea that it is needed to handle the problem of indeterminacy. However, it should be clear, if strong normativism is to help with this problem it must be understood as a *metasemantic* claim – i.e. as a claim about the facts in virtue of which ‘dog’ has a certain meaning. And that, very clearly, is a metaphysical claim about the property of meaning DOG by ‘dog’, not just a claim about the concept.

Consider Gibbard’s discussion of Horwich’s version of the use theory of meaning. According to Horwich the meaning of an expression consists in a basic ‘acceptance property’: the property that one is prone (under certain conditions) to accept certain sentences containing the word. For instance, the fact that ‘That is red’ is accepted for red things, serves to determine the meaning of the component terms. To account for error, Horwich speaks of ‘ideal laws’ governing linguistic usage such that what determines meaning is acceptance properties under ideal conditions. Gibbard is sympathetic to Horwich’s approach but argues that the project is not wholly successful since it fails to separate out which sentences are meaning-determining and which substantive, thereby leaving room for indeterminacy – at least when it comes to theoretical terms, such as ‘mass’ or ‘phlogiston’. To illustrate this Gibbard considers a subject, Newtonia, who accepts four central claims of Newtonian theory. Given this acceptance pattern, Gibbard suggests, there are several possible interpretations of Newtonia’s term ‘mass’, and Horwich’s theory does not tell us which is the correct interpretation, since it does not tell us “how to go from her performance, the actual dispositions to deploy her terms ‘m’ and ‘p’, to an ideal explanatory model of these dispositions” (108). The remedy, Gibbard suggests, is to appeal to oughts: “Meanings explain
usage, but if the causal-explanatory demands on meanings leave slack in determining what terms mean, perhaps the oughts of the matter remove some of the slack” (p. 110).

How could normativism help remove this slack? Not simply by saying that meaning statements imply ought statements. That is, let us assume that there are two competing interpretations of Newtonia’s term ‘mass’, as rest mass or relativistic mass:

(i) The term ‘mass’ as used by Newtonia means $\text{MASS}_{\text{rest}}$

(ii) The term ‘mass’ as used by Newtonia means $\text{MASS}_{\text{rel}}$

Following Gibbard, the two statements imply different ought statements. For instance, (i) implies that Newtonia ought to accept ‘mass is constant’ whereas (ii) does not. Let us assume that all the implied semantic ‘ought’-statements have been sorted out, characterizing the two distinct meanings, $\text{MASS}_{\text{rest}}$ and $\text{MASS}_{\text{rel}}$. To solve the indeterminacy problem, clearly, we need to do more work: We need to appeal to facts about Newtonia in virtue of which her word means one thing rather than the other. That is, we need to do metasemantics. Now, as noted above, normativism could be construed as a metasemantic thesis, i.e. as a version of MD-normativism. Thus, one might try to solve the problem of indeterminacy by appealing to the idea that meaning facts are determined by facts about rule guidance. Understood that way, normativism tells us that Newtonia means $\text{MASS}_{\text{rest}}$ if she is guided by one set of oughts, and $\text{MASS}_{\text{rel}}$ if she is guided by another set of oughts. However, there are reasons to suspect that Gibbard does not wish to adopt MD-normativism.

First, Gibbard explicitly denies that we are to think of the speaker as being guided by the relevant oughts, in the sense that she thinks she ought to do X and as a consequence does X (p. 117). He denies this for good reason, since his normativism is meant to apply both to linguistic meaning and to mental content. As argued in Wikforss & Glüer 2009b, it cannot be that mental content is determined in virtue of the subject’s following certain rules, on pain of a regress of contents, since being guided by a rule requires having
‘ought-thoughts’ of the sort Gibbard mentions. If, therefore, Gibbard were to construe strong normativism as a form of guidance normativism, he would not only have to abandon the claim that the strong thesis is not about the property, he would also have to abandon the claim that normativism applies both to meaning and to mental content.

Second, Gibbard stresses that he does not wish to offer an alternative to Horwich’s substantive theory of meaning (p. 22), but is merely interested in the meaning of ‘meaning’. Since MD-normativism is a substantive theory this cannot be how Gibbard’s strong thesis is to be understood. If so, however, the thesis does not help with the problem of indeterminacy – only a substantive theory could do that. No doubt, Gibbard does have some things to say when it comes to the substantive theory of meaning. However, what he says merely points to a version of dispositionalism. For instance, having argued that asserting that Pierre’s word ‘chien’ means DOG involves asserting that a certain ought pattern holds, Gibbard raises the question in virtue of what this ought pattern obtains. He suggests that it obtains “in virtue of patterns in Pierre’s usage tendencies, his dispositions to accept sentences in his own language in various circumstances” (p. 138). Meaning DOG by ‘chien’, Gibbard continues, is constituted by a certain “pattern of dispositions”. If other versions of dispositionalism have troubles with the problem of indeterminacy (and related problems, such as the problem of error) then so does Gibbard’s version. Adding the whole metatheory about the meaning of ‘meaning’ being normative changes nothing when it comes to these metasemantic challenges.

In the end, I think, Gibbard is not really interested in the problem of indeterminacy. Rather, his project is a very different one: That of explaining the nature of metasemantic disagreements. For instance, in the discussion of Horwich he suggests that his aim is not to offer a substantive theory of meaning but merely to offer a metatheory of meaning which “explains what’s at issue between Horwich’s theory and alternative theories
of meaning, and what’s at issue in the matter that Horwich’s substantive theory doesn’t resolve” (p. 22). This is also quite explicit in chapter 2, where Gibbard discusses two philosophers, Tyler and Jerry, who disagree in their metasemantic theories: While Jerry thinks that meaning is determined by intrinsic properties of the individual’s brain, Tyler thinks meaning is determined by the dispositions of the larger community to use ‘+’. This disagreement, Gibbard stresses, is a disagreement “about what the natural facts are in virtue of which statements mean what they do” (p. 40). The challenge, according to Gibbard, is to explain how Jerry and Tyler can mean the same thing by ‘means’, make no mistake of usage, and still disagree with one another, and the answer is that their disagreement is normative: “Jerry and Tyler disagree on the normative standards that settle which sentences, in a person’s own language, the person ought to accept” (p. 44).

In the end, therefore, Gibbard’s main argument in support of meaning normativism seems to be that it is required if we are to be able to explain disagreements in metasemantics: Just as two philosophers might agree on all the natural properties of an action and disagree on whether the action is morally right, two philosophers might agree on all the natural properties of a speaker and disagree on what she means by ‘plus’. However, I do not find this a very compelling reason to adopt Gibbard’s meaning normativism. It is not difficult to think of philosophical disagreements where philosophers agree on all the natural properties, but disagree philosophically: Is action A free if it is caused? Is person S the same as ten years ago? Does a zombie have pain? Are we to conclude that every such disagreement is a disagreement over oughts? Surely not. These are substantive disagreements, concerning difficult metaphysical issues. And so is the disagreement between Jerry and Tyler. Of course, the metasemantic disagreement, in their case, implies a disagreement in semantics: For instance, they will disagree on the truth-conditions of Ursula’s statements involving ‘plus’.
But, again, we have been given no reason to think that statements about truth conditions imply ought-statements.

In sum, I do not think that Gibbard has sufficiently motivated extending normativism to semantics: The simple argument does not succeed, and the more substantial arguments provided, in particular the argument from indeterminacy, are problematic. If I am right about that, then the territorial aspirations of metaethics are unwarranted: It is not the case that a central topic in the philosophy of language becomes a part of metaethics. No doubt, to the extent that metaethics is concerned with the meaning of value terms such as ‘good’ and ‘right’, metaethics is concerned with semantics, but this is rather philosophy of language grabbing territory from metaethics.

References


---

i All page references are to Gibbard 2012.

ii Glüer & Wikforss 2009a.

iii We note that although most people take conceptual necessity to imply metaphysical necessity this can be questioned – as Gibbard seems to do. I shall return to this below.

iv Gibbard seems to agree with all this. For instance, he appeals to the idea that the relevant ‘ought’ is a primitive ‘Ewing ought’: an ‘ought’ that follows invariably, independently of the subject’s desires and further considerations (pp. 13-14). And he stresses that the relevant ‘oughts’ must flow from meaning itself, and not from epistemic or other considerations (pp. 114-115).

v Following Stalnaker 1997, a *semantic* theory (or what he calls ‘descriptive semantics’) is a theory that provides the semantic content of our linguistic expressions, whereas a *metasemantic* theory tells us something about the facts in virtue of which an expression has a given semantic content.

vi The distinction between two types of normativism applies equally at the level of mental content: Content normativism might be understood either as the thesis that content ‘engenders’ certain oughts (‘CE-normativism’), or as the metasemantic thesis that mental content is determined by certain norms being in force for S (‘CD-normativism’). See Glüer & Wikforss 2009a and 2009b.

vii See Glüer & Wikforss 2009a and 2015.


ix For a discussion of this point see Glüer & Wikforss 2015.

x For instance, there are notorious difficulties singling out the inferences that are pure meaning inferences from the ones that are not, and adding oughts into the picture does not
help with this problem. What if I do not infer ‘Bob has a mammal’ from ‘Bob has a dog’ since I harbor the suspicion that all dogs are robots controlled from Mars? Must it be concluded that I have violated a semantic ‘ought’? See Williamson 2007, chapter 4, for a criticism of inferential role semantics along these lines.

On p. 117 Gibbard refers to our 2009b paper and seems to accept our argument although he suggests that there is a notion of immediate inference that is not subject to the regress. I do not think that this avoids the regress problem, since I do not think that immediate inferences (in Gibbard’s sense) involve rule guidance, but I shall have to leave the issue be here.

It is important to bear in mind that by ‘metatheatery’ Gibbard means a theory about the meaning of ‘meaning’, not metasemantics.