Mark Schroeder
University of Southern California
March 15, 2009

how to be an expressivist about truth

In this paper I explore why one might hope to, and how to begin to, develop an expressivist account of truth – that is, a semantics for 'true' and 'false' within an expressivist framework. I do so for a few reasons: because certain features of deflationism seem to me to require some sort of nondescriptivist semantics, because of all nondescriptivist semantic frameworks which are capable of yielding definite predictions rather than consisting merely of hand-waving, expressivism is that with which I am most familiar, and because I believe that certain problems about truth and particularly about paradox seem to me to look different, when seen through the lens of an expressivist theory. I don't mean to defend such a theory in this paper, and indeed I have cast doubts on the ultimate prospects of the framework I will be employing here elsewhere.[1] But I do think that seeing what an expressivist theory of truth would look like helps to shed light on both expressivism and on truth.

1.1 intersubstitutability

I'll first motivate the expressivist treatment of truth by consideration of two different issues about truth, one having to do with intersubstitutability, and one having to do with the psychological attitude of rejection. Intersubstitutability first; here the problems with truth start with the observation that on the assumption that the meaning of a sentence S is that P, 'S is true' seems, at least at first glance, to be substitutable pretty much anywhere for 'P', and conversely. On the face of it, this observation is precisely what we need in order to get the full usefulness out of the truth predicate as a 'device of generalization',[2] but it is also sufficient to raise the spectre of paradox, in connection with sentences like Liar:

Liar: Liar is not true.

[1] Schroeder [2008], especially chapter 12.

[2] See, for example, Gupta [2005] and Field [2008], especially chapter 13. I won't be concerned in this paper to defend full intersubstitutability; I'll simply be concerned with what it would take to capture it.

By stipulation, Liar says of itself that it is not true – so it seems indisputable that the meaning of Liar is that Liar is not true. So 'Liar is true' should be substitutable everywhere for 'Liar is not true'. Now all it takes to raise a problem is to assume that Liar is true just in case Liar is true. Once we assume that much, intersubstitutability leads us to the conclusion that Liar is true just in case Liar is not true, which reeks of paradox.[3]

Some theorists about truth, who are persuaded of the force of the reasons why we should want full intersubstitutability, argue that any solution to the paradox *must* work by finding a reading of the conditional, 'Liar is true just in case Liar is not true', which we can comfortably accept without that allowing us to derive an outright contradiction, as in 'Liar is true and Liar is not true'.[4] These theorists argue, in effect, that we can't have full intersubstitutability without this consequence.

There are two ways in which this reasoning can work. The first proceeds, as we did above, by assuming that Liar is true just in case Liar is true, and then to apply intersubstitutability to this biconditional. However, this reasoning is problematic. Even without this assumption, intersubstitutability means that if we accept 'Liar is true', we are committed to accepting 'Liar is not true', and if we accept 'Liar is not true', we are committed to accepting 'Liar is true'. So the obvious conclusion is that the only rational course is to accept neither. And if you are committed to neither accepting 'Liar is true' nor 'Liar is not true', you shouldn't accept 'Liar is true or Liar is not true', either – for that's what allows the reasoning by cases to get started, which leads to a contradiction. But on its material conditional reading, 'Liar is true just in case Liar is true' is *equivalent* to 'Liar is true or Liar is not true'. So of course you can't accept that. So unless there is some different reading of the biconditional on which it does not entail the material biconditional, you simply shouldn't accept 'Liar is true just in case Liar is true', to begin with. So this line of reasoning provides no real reason to think that you must accept the paradox-inducing biconditional, 'Liar is true just in case Liar is not true'.

The line of reasoning that remains, then, argues that we *need* some sort of biconditional in order to be able to articulate the force of intersubstitutability. Theorists who pursue this line hold that we need to have some sort of conditional in English which is able to articulate intersubstitutability, of the form:

T-schema If S means that P, then S is true just in case P.

[3] On the assumption that Liar is true, it leads us to the conclusion that Liar is not true, and on the assumption that Liar is not true, it leads us to the conclusion that Liar is true. So reasoning by cases forces us to conclude that Liar is true and Liar is not true – an outright contradiction.
[4] See, for example, Gupta [2005] and Field [2008].

On this view, the reason why ‘S is true’ and ‘P’ are intersubstitutable conditional on the assumption that S means that P, is that from ‘S means that P’ and T-schema, we can derive the biconditional ‘S is true just in case P’, and it is that biconditional which licenses intersubstitutability. Of course, this can’t be a material biconditional, nor indeed any conditional which entails the material biconditional, because one instance of T-schema is the following:

T-Liar If Liar means that Liar is not true, then Liar is true just in case Liar is not true.

The assumption of this line of reasoning, is that to articulate intersubstitutability, we need to be able to somehow *lexicalize* the way that, if you accept that the meaning of S is that P, accepting ‘P’ commits you to accepting ‘S is true’ and conversely – and specifically that we need to do so by using a *conditional*, which would lead us to endorse every instance of some conditional like T-schema (on the proper interpretation of the conditional).

This particular line of reasoning appears to have launched quite a large research project in the technical literature on truth – the search for a conditional weaker than the material conditional, so that we can accept ‘P iff ~P’ without being led to accept ‘P&~P’. It plays a particularly large role in the work of, for example, Anil Gupta and Hartry Field. But I’m skeptical. What I want to understand better, is whether there isn’t a better way of capturing full intersubstitutability than actually *endorsing* every instance of T-schema. As we’ll see later on, an expressivist semantic framework is suited to explain intersubstitutability without requiring us to be able to endorse every instance of anything like T-schema. The special status of T-schema won’t be that we are committed to *accepting* all of its instances, but rather that we are committed to not *denying* any of them. All of this is compatible with the idea that we may *reject* some instances. So it is to rejection that I now turn.

1.2 rejection

If you don’t believe that Liar is true, and you don’t believe that Liar is not true, what are you to do? If there is no proposition expressed by Liar, then the answer may seem to be simple: there is nothing *to* believe. But that answer is no good. If there is no such thing as the proposition that Liar is true to believe, and no such thing as the proposition that Liar is not true to believe, then there must not be any such thing as the proposition that Liar is true *and* Liar is not true to believe, either, and so you are safe from believing any contradiction, even if you accept and are willing to assert and act on both 'Liar is true' and 'Liar is not true'. If only things were so easy!

The thesis that Liar does not express a proposition also runs into the problem that many liar-paradoxical sentences are only contingently paradoxical. Take, for example, the following sentence:

Contingent Liar The only sentence written on the whiteboard in Hartry Field's office is not true.

If it so turns out that Contingent Liar is the only sentence written on the whiteboard in Hartry Field's office, then this sentence is liar-paradoxical; if not, then not. But it hardly seem plausible that there should be such a thing as the proposition that the only sentence written on the whiteboard in Hartry Field's office is not true, which can even be expressed on the whiteboard in Hartry Field's office, but which goes out of existence as soon as everything else is erased from the whiteboard. Indeed, there is direct evidence against this; for even if this is the only thing written on Hartry Field's whiteboard, people who do not know or suspect this may wonder whether the only sentence written on the whiteboard in Hatry Field's office is not true, and may believe or disbelieve that the only sentence written on the whiteboard in Hatry Field's office is not true. If this is something that they believe, then any reason to think that belief is a relation to propositions will carry over to this case. So if propositions are the objects of belief, we should conclude that there really are paradoxical propositions, in cases like this one.

So if there really is a proposition expressed by the liar, what attitude should you take toward it? Believing it leads to paradox, denying it (in the sense of believing its negation) also leads to paradox. So what should you do? The obvious answer is: you should *reject* this proposition. If there is a third attitude of rejection that it makes sense to have toward a proposition – which is an alternative to both belief and belief in its negation, then that is the right one to take to the proposition expressed by Liar, and in general toward other such paradoxical propositions. I will have more to say in a little bit about what sort of attitude rejection might be, but for now it suffices to observe that given that liar-paradoxical sentences really do express propositions, it would be very nice to hope that there is such an attitude which it makes sense to take toward those propositions – an alternative to believing them and believing their negations, which doesn't amount to merely being unsure about them.

So far, so conventional – many authors have advocated rejection as an appropriate response to Liar.[5] But just as in the last section we saw that some philosophers have thought it imperative to be able to articulate intersubstitutability by being able to lexically express it with a conditional, some philosophers have thought it important to be able to lexicalize rejection, by being able to articulate it with a special meaning of 'not'. Take, for example, the recent approach of Mark Richard [2008], who believes that in addition to the ordinary, truth-conditional 'not', which allows someone who believes that Liar is not true to express her view by saying, 'Liar is not true', there is a special sense of 'not' which allows someone who *rejects* the proposition[6] that Liar is true by saying, 'Liar is not* true'.

Lexicalizing rejection, however, is deeply problematic. Once we lexicalize rejection (as Richard recognizes), that gives rise to paradoxes of revenge, by providing us with ways to formulate new paradoxical sentences, whose contents it is just as paradoxical to reject as to either believe or believe their negations. For example:

Liar's Revenge Liar's Revenge is not true or Liar's Revenge is not* true.

If you accept Liar's Revenge and understand what it means, then by intersubstitutability you are committed to accepting that Liar's Revenge is true – which in turn allows you to infer by disjunctive syllogism that Liar's Revenge is not* true – i.e., which commits you to rejecting Liar's Revenge. If you accept the negation of Liar's revenge, then by intersubstitutability you are committed to accepting that Liar's Revenge is not true, and hence to accepting Liar's Revenge – a contradiction. And if you reject Liar's Revenge, then you accept 'Liar's Revenge is not* true' – which commits you to accepting Liar's Revenge. So there is no consistent attitude, out of the trio of acceptance, rejection, and acceptance of the negation, for you to take toward Liar's Revenge. So if Liar's Revenge really expresses a proposition – and analogous reasoning to the foregoing will lead to the conclusion that it does – that leaves the question of what to do with this proposition unanswered, unless we postulate yet a *further* kind of attitude, rejection-prime, to have toward this proposition.

Richard [2008] (for example) happily marches off of this cliff, but I'm not so sure; just as the problem about intersubstitutability arose from the idea that we had to be able to lexicalize the commitment relationship involved in intersubstitutability with a conditional, this problem arises from the idea that we

[5] For recent discussions, see particularly Field [2008] and Richard [2008].

[6] Richard wouldn't use the word 'proposition' here, but I'll use it for consistency. Richard also prefers 'denial' as a name for rejection, although he goes back and forth.

have to be able to – or even can – lexicalize rejection, in the sense that there is a sentence which it makes sense to endorse just in case you reject the proposition that P.

1.3 two observations, and the link to expressivism

There are two important observations to be made, here. The first is that both the question of intersubstitutability and the idea that liar-paradoxical sentences are ones that it makes sense to reject are theses *about the rationality of mental states*. Intersubstitutability says that you are committed to having the same attitude to the propositions expressed by 'P' and by 'S is true', conditionally on accepting or at least not denying that the meaning of S is that P. And the idea about rejection is that rejection is the attitude that it *makes rational sense* to have toward liar-paradoxical propositions. If there is a semantic framework which ought to be in the best position to be able to articulate these ideas, it would be *expressivism*, whose central idea is that a semantic theory should work by associating each sentence, 'P', with what it is to think that P, and whose treatment of logical inference works directly by articulating rational commitment relationships between mental states.

The second important observation to be made, here, is that both the issues about intersubstitutability and the issues about rejection turn on the question of what sorts of things we should be able to *lexicalize*, and how. The problem about intersubstitutability arises from the idea that if each of us is in the position of being rationally committed to having the same attitude toward 'P' and toward 'S is true', conditionally on accepting (or at least not denying) that the meaning of S is that P, then there must be some sentence, T, which *expresses* that state of conditional commitment, which we are all committed to accepting. Similarly, the problem about rejection arises from the idea that if you reject some proposition, that P, there must be some sentence, R, which *expresses* that state of rejection.

Within an expressivist framework, it is easy to question both of these assumptions – expressivism is founded on the importance of the distinction between *expressing* a mental state, and reporting that you are in it, and much contemporary work on expressivism has provided very strong reasons to think that not just any mental state is one that it is possible to express with a sentence.[7] To say that we are all in a state of being rationally committed to having the same attitude toward 'P' and toward 'S is true', conditionally on

[7] On the significance of the distinction between expressing and reporting, see chapter 2 of Schroeder [2008]. For the importance of the idea that not all mental states are ones that it is possible to express, see Gibbard [2003], especially chapter 4. In Gibbard's framework, all and only mental states with which it is possible to *disagree* can be expressed by indicative sentences; for justification of a yet-more-restrictive account of which states can be expressed by sentences, see chapters 3-8 of Schroeder [2008].

accepting (or at least not denying) that the meaning of S is that P is to *report* a state that we are in, and quite different from expressing that state. Indeed, there may be no sentence at all – not even a possible sentence – which expresses that state, in which case there would be no way of putting the T-biconditional that we would all be committed to accepting. Similarly, to say that you reject the proposition expressed by Liar is to *report* on your mental state, and is quite different from expressing that state. Indeed, there may be no sentence at all which expresses that state, in which case there would be no way of framing Liar's Revenge, which would require some response other than rejection.[8]

At any rate, these observations make me interested in the hypothesis that the state of conditional commitment described by full intersubstitutability and the attitude of rejection are states which can be reported, but which cannot be expressed by natural-language sentences, and in the idea that this can be fruitfully explored within an expressivist framework. In the next section I'll briefly motivate a certain abstract perspective on both of these issues which is neutral between different detailed expressivist frameworks, and then in the remainder of the paper I'll illustrate how these ideas can be made good on in a rigorous way within the semantic framework of biforcated attitude semantics, developed in Schroeder [2008].

1.4 a simple framework: commitment theory

Start with the idea that there are three 'committed' attitudes which it is possible to take to a proposition: *acceptance*, *rejection* and *denial*, and assume both that denying *p* is just accepting *p*'s negation, and that rejecting *p* and rejecting ~*p* are the same state. Assume that each pair of these three attitudes toward *p* are rationally inconsistent, in the sense that Allan Gibbard [2003] calls *disagreement*: for any two thinkers, if they bear two different of these three attitudes toward the same proposition, then they disagree with one another. Let us say, further, that if you are in a state of mind which disagrees with two of these three attitudes toward a proposition *p*, then you are *committed* to the third attitude toward *p*.

This framework allows us to think of the strong Kleene tables as *commitment* tables, rather than truth-tables, in the sense that given the attitudes that a speaker has toward some propositions, they tell us what commitment that speaker is committed to having toward other propositions. For example, take the case of negation:

[8] Compare, for example, the attitude of *doubt*. There need be no sentence, 'Q', such that doubting that P is equivalent to thinking that Q. The same thing may go for rejection.

P	~P
A	D
R	R
D	A

The first line of this commitment table tells us that if you accept P, then you are committed to denying ~P, in the sense that your state of mind disagrees with both that of rejecting ~P and that of accepting ~P. These facts follow from our assumptions: rejecting ~P is the same state as rejecting P, and we assumed that accepting P disagrees with rejecting P. So accepting P disagrees with rejecting ~P. We also assumed that accepting ~P is the same state as denying P, and we assumed that accepting P disagrees with denying P. So accepting P disagrees with accepting ~P. Hence, by the definition of commitment, someone who accepts P is committed to denying ~P, as the first line of the commitment table indicates. Similar observations go for the other lines – the commitment table is simply a clear and clean way of articulating these assumptions and exhibiting their properties.

Note that this is a *commitment* table, not a *truth* table. The table tells us nothing about the semantic status of P or of ~P; it only tells us which combinations of attitudes toward P and ~P avoid the kind of inconsistency involved with disagreeing with oneself. But a similar table can be constructed for '&':

P	Q	P&Q
A	A	A
A	R	R
A	D	D
R	A	R
R	R	R
R	D	D
D	A	D
D	R	D
D	D	D

Again, this table tells us what attitude someone is committed to having toward P&Q, on the basis of the attitudes she has toward P and toward Q. This time, the properties of the table cannot be derived solely from our assumptions; but they do follow from plausible assumptions about disagreement and conjunction – for example, all of the D lines for P&Q follow from the assumption that denying either P or Q disagrees with either accepting or rejecting P&Q. Again, this is *not* a truth table. It is simply a *commitment* table telling us which combinations of attitudes avoid self-disagreement.

The assumption that these are the right commitment tables for '~' and '&' is sufficient to guarantee two important results, on the basis of familiar facts about the strong Kleene tables. The first of these is that every theorem of classical logic is undeniable, in the sense that there is no rationally consistent set of attitudes which involves denying it. The second of these is that *modus ponens* using the material conditional ('P⊃Q' defined to mean '~(P&~Q)') takes you from propositions that you accept only to propositions that you are already committed to accepting – so it preserves commitment.[9] These facts are not quite enough for it to follow that all classically valid rules of inference preserve commitment, because arbitrary classical theorems follow from arbitrary assumptions, and under these assumptions, one need not be committed to accepting arbitrary classical theorems – only to not denying them. But they are sufficient to explain why all classically valid rules are commitment-preserving for anyone who accepts 'P∨~P' for each atom 'P'.

So any theory which is in a position to explain these commitment tables is in a position to explain all of these things about the relationship between what thinkers are committed to and classical logic. A fruitful way to think about the aspirations of an expressivist semantic theory is through the lens of commitment theory, whose main ideas I have been sketching here. An expressivist semantics will aspire to say, for each sentence 'P', what it is to accept that P, what it is to reject that P, and what it is to deny that P, on the basis of compositional rules which have the following consequence: that the rule for '~' predicts the commitment table for '~' and the rule for '&' predicts the commitment table for '&'.

A similar aspiration would go for an expressivist account of truth: there would be some commitment table for 'TRUE', and it would be the aspiration of the expressivist theory to say, for each sentence 'P' involving 'TRUE', what it is to think that P, in such a way that the commitment table for 'TRUE' would be predicted by that account. The commitment table for a sentential truth predicate 'TRUE' should look like this:

[9] The first of these facts follows from the result that LP, the logic resulting from the choice of both 'A' and 'R' as 'designated' values, has the same theorems as classical logic. The second follows from the result that K_3, the logic resulting from choice of only 'A' as the 'designated' value in the strong Kleene scheme, validates *modus ponens*. See, for example, Avron [1993] for proofs and discussion.

P	S means that P	TRUE(S)
A	A	A
R	A	R
D	A	D
A	R	A
R	R	R
D	R	D
A	D	
R	D	
D	D	

Again, note that this is not a truth table. What it tells us, is that for any sentence P, someone who either accepts or rejects that S means that P is committed to having the same attitudes toward ‘P’ and toward ‘S is true’. In other words, it articulates the idea of full intersubstitutability.

As we should expect, two very nice things follow from this commitment table for ‘TRUE’, along with the commitment tables for ‘~’ and ‘&’. The first is that if we use the material conditional, then T-Schema is rationally *undeniable*:

T-Schema $\text{MEANS}(S,\text{THAT}(P))\supset(\text{TRUE}(S)\equiv P)$

So even though we needn’t *accept* every instance of T-Schema, on this picture, that doesn’t mean that T-Schema is irrelevant; on the contrary, because of the commitment table for ‘TRUE’, T-Schema is exceptionlessly undeniable – that is, it has the very same status as the theorems of classical logic. Nothing rationally commits us to accepting every instance of them, either, if we reject their atoms, and in that respect T-Schema still has a privileged status.[10] The second very nice thing which follows from the commitment tables, is that if you accept that Liar means that Liar is not true, you are committed to rejecting both Liar and T-Liar, the corresponding instance of T-Schema. This is as should be expected.

In this section I’ve sketched in a rudimentary and quick way some of the basic ideas of commitment theory, which gives us a fruitful way of talking about the rational relationships between the different attitudes that we might take toward sentences, or toward the propositions which they express. Commitment theory gives us all of the resources that we need, in order to be able to articulate both the thesis of full intersubstitutability and the diagnosis that the right response to Liar is to reject it, as well as

[10] Compare Tappenden [1993], who suggests that T-sentences are guaranteed by the meaning of ‘true’ not to be false; the idea being developed here can be thought of as a paracomplete analogue of Tappenden’s theory.

to explain why T-schema exerts 'pull' and seems to play an important role in every single instance, even though not every instance is acceptable.[11,12] In part 2 I'll connect commitment theory to expressivism, by showing how an expressivist framework can both predict each of these commitment tables and explain why there is no way of lexicalizing rejection – of forming a sentence which expresses, rather than reports, the state of rejecting some proposition.

2.1 expressivist semantics

As I noted above, an expressivist semantics works by associating each sentence, 'P', with what it is to think that P. I think of it as a kind of assertability-conditional semantics, where instead of the assertability conditions of a sentence being some features of the world, the assertability conditions are always that the speaker be in the appropriate mental state. This corresponds to the idea that if someone who falsely thinks that P asserts 'P', the mistake that she makes is one about the world, rather than in her semantic grasp of the language. The assertability conditions set out by an expressivist theory are the ones which track the conditions under which a sincere speaker does not make a mistake in her semantic grasp of the language.

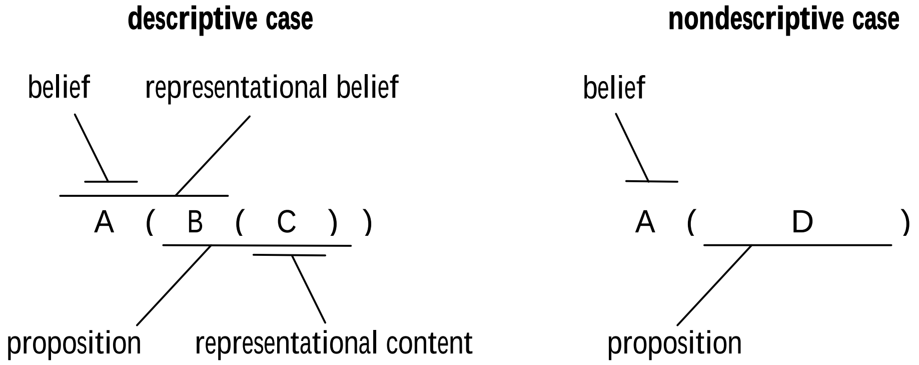
So conceived, expressivism is a kind of nondescriptivist semantic theory. It is sometimes said that nondescriptivists like expressivists do not really believe in propositions, or have only a deflationary conception of propositions – at least for the special domain of which they seek to provide a distinctively nondescriptivist account. But I prefer not to think of things this way. Instead, I prefer to think of expressivism as divorcing two sets of theoretical roles for propositions. One set of theoretical roles for propositions is to be the bearers of truth and falsity and objects of attitudes like belief, desire, and assertion. But propositions are also often said to play a role in carving up the world at its joints, to be connected in some way to metaphysical commitments, and to be the appropriate objects of excluded middle. My suggestion is that we think of expressivism as divorcing these two theoretical roles for propositions. I'll reserve the word 'proposition' for whatever entities are the objects of the attitudes and the bearers of truth and falsity, and I'll use the word 'representational content' for whatever entities carve

[11] Compare, for example, Eklund [2002] and Tappenden [1993] on the unrestricted 'pull' of T-Schema.

[12] Commitment theory contrasts fruitfully, I think, with the constrained approach of standard three-valued logic, which washes out much of the information available from the strong Kleene tables by forcing a false choice between which values are to count as 'designated'. If the only information that we can glean from the tables is fixed by the choice of designated values, then we would have to choose between whether the tables allow us to capture the theorems of classical logic, and whether they allows us to capture the idea that *modus ponens* is valid – the former but not the latter holds for LP, in which 'A' and 'R' are both 'designated', and the latter but not the former holds for K_3, in which only 'A' is designated. In commitment theory, in contrast, we can say something informative about both the theorems of classical logic and the validity of *modus ponens*: the theorems are *undeniable*, and *modus ponens preserves commitment*.

up the world at its joints, are connected with metaphysical commitments, and are the appropriate objects of excluded middle. The main idea of expressivism, then, and indeed of any nondescriptivist framework which accepts this divorce, is that every indicative sentence can be associated with some proposition, but some cannot be associated with a corresponding representational content. The sentences which cannot be associated with any representational content are the *nondescriptive* sentences of the language. Expressivism about truth is the idea that sentences involving the word 'true' are nondescriptive.

Contrasting with nondescriptive sentences are ordinary descriptive sentences, which correspond to both a proposition and a representational content. For example, if part of the structure of reality, when the world is carved up at its joints, is that green is a way that grass may or may not be, then 'grass is green' will correspond to both the proposition that grass is green and the representational content of grass's being green. It is natural to think of the ordinary descriptive belief that grass is green as a single state which can be alternately thought of as a relation to *either* of these objects. This is possible, if ordinary descriptive belief has the sort of structure illustrated by the following two diagrams:



If all belief involves an attitude, A, toward propositions, and some propositions are themselves constituted by a relation or property, B, toward or of representational contents, then so long as we stick to the descriptive case, there will be a way of describing the very same state either as a relation toward propositions, or as a relation toward representational contents. If the descriptive case is the paradigm case, and this is really the structure of belief, then it would be no wonder that we could easily have confused propositions with representational contents. But if this is really the structure of belief, then there may also be a second, nondescriptive, case, in which the proposition does not itself involve a relation to any representational content. This is a useful way of understanding the basic structure of belief according to biforcated attitude semantics, the expressivist framework developed in Schroeder [2008].

The basic building block of biforcated attitude semantics is an attitude toward properties that I call 'being for', and which I assume to have the property that two states of being for disagree with one another, in Gibbard's sense, just in case their objects are inconsistent properties. I think of the state of being for as one which, when other things are equal, makes one who is in it come to acquire the property that is its object – but nothing about the basic structure of biforcated attitude semantics turns on this. The generic attitude of belief – the one that takes propositions, rather than representational contents – is constructed out of a pair of states of being for, one of whose objects is strictly stronger than the other.[13] Such a state I call a *biforcated attitude*, for obvious reasons. Propositions, in this framework, are just pairs of properties,[14] one of which is strictly stronger than the other, and the *belief* relation is just the relation of being for each member of the pair.

Representational belief, in biforcated attitude semantics, is assumed to be a special case of a biforcated attitude. To see which case this is, we need to appeal to the relation of *proceeding as if*, which I take to be a relation between agents and representational contents. Intuitively, an agent proceeds as if some representational content just in case she takes it as settled in deciding what to do. The only assumption about proceeding as if which I require, however, is that proceeding as if *p* is incompatible with proceeding as if ~*p*. It follows from this that for each representational content, *p*, there is a pair of properties, consisting of the property of proceeding as if *p* and the property of not proceeding as if ~*p*, and that this pair of properties is a proposition. Being for each property in such a pair is what it is to have an ordinary representational belief whose object is *p*. So on this picture, ordinary representational belief is just one special case of a broader class of states, and propositions associated with representational contents are just one special case of a broader class of propositions.

2.2 connectives, rejection, and logic in biforcated attitude semantics

It is easy to define both the connectives and rejection in biforcated attitude semantics. For any proposition, P, there is a pair of properties consisting of the negations of each property in P. It is easy to see that this pair also has one member which is strictly stronger, so it is a proposition. We define it to be the negation of P. Moreover, for any two propositions, P and Q, there is a pair of properties consisting of the conjunction of the stronger members of each of P and Q, and the conjunction of the weaker members

[13] The assumption that one is *strictly* stronger is new here; I did not assume this in Schroeder [2008].

[14] In Schroeder [2008] I used 'proposition' as a word for what I here call representational contents, and only briefly in chapter 11, raised the question of whether it

of each of P and Q. It is easy to see that this pair also has one member which is strictly stronger, and hence that it is a proposition. We define it to be the conjunction of P and Q.

To connect up biforcated attitude semantics with commitment theory, we may identify acceptance with belief, and denial with belief in the negation. Rejection is also easy to define; to reject P is to be for the weaker member of each of P and ~P. It is straightforward to derive the properties we assumed in section 1.4 about acceptance, rejection, and denial, from our assumption that two states of being for disagree with one another just in case their objects are inconsistent properties. It is also straightforward to observe that no state of rejection is identical to any biforcated attitude, and hence that no state of rejection is identical to any belief. So if all sentences express biforcated attitudes, there will be no sentence which expresses a state of rejection. That is, rejection cannot be lexicalized in biforcated attitude semantics – making good on one of the observations with which we began.

These assumptions also suffice, straightforwardly, to predict the commitment tables for '&' and '~', which the reader may either verify for herself or consult chapter eight of Schroeder [2008]. Consequently, biforcated attitude semantics constitutes a framework with flexibility to allow for both descriptive and nondescriptive atomic sentences, either of which incorporates equally well into a single, unified picture of the mental states expressed by complex sentences, and which predicts the rational relationships among all of these sentences, as well as their relationship to those sentences' intuitive logical properties, as spelled out in section 1.4. This, I believe, is the basic thing that we should expect from any viable nondescriptivist semantics, including an expressivist one.

2.3 predicates of propositions in biforcated attitude semantics

Because biforcated attitude semantics treats propositions as pairs of properties, it has a candidate available to be the subject of predicates like 'believes that', 'said that', 'means that' and 'it is true that'. This is an important improvement over approaches which allow only 'deflationary' talk about propositions, because it allows for straightforward quantification into the propositional argument place. We may easily accommodate this by treating 'believes', 'said', 'means' and 'it is true' as predicates of propositions, and treating 'that' as an operator which takes a sentential complement and denotes a proposition (relative to an assignment to the unbound variables in the complement, if there are any). This is, of course, not an unfamiliar idea, save that in the framework of biforcated attitude semantics, we have a different (and surprising) underlying picture of what propositions are like.

In what follows, I'll assume that the first three of these are descriptive predicates, and that 'true' is a nondescriptive predicate. What that means, again, in biforcated attitude semantics, is that atomic sentences formed using the first three predicates express what I've here called representational beliefs. Since belief is a relation between an agent and a pair of properties one of which is stronger than the other constituted by the agent being for each property in the pair, the proposition that *x* believes *y* will be the pair of properties consisting of the property of proceeding as if *x* is for each of the properties in *y*, and the property of not proceeding as if *x* is not for each of the properties in *y*. Similarly, assuming that we want 'S means that P' to report, essentially, that the proposition expressed by S is the proposition that P, the proposition that *x* means *y* will be the pair of properties consisting of the property of proceeding as if the proposition expressed by *x* is identical to *y* and the property of not proceeding as if the proposition expressed by *x* is not identical to *y*. Similar moves suffice to deal with 'said'.[15]

If 'true' is to be a nondescriptive predicate of propositions, then atomic sentences involving 'true' may express any pair of properties of which one is strictly stronger. But only some choices of such a pair allow us to predict the commitment tables for 'true'. Here is one: for any proposition, *y*, consider the pair of properties consisting of the property of instantiating the stronger member of *y* and the property of instantiating the weaker member of *y*. Since the stronger member of *y* is strictly stronger than its weaker member, the property of instantiating the stronger member of *y* is strictly stronger than the property of instantiating the weaker member of *y*. So this pair is a proposition. Let it be the proposition that *y* is true.

This account allows us to predict the following commitment table for propositional truth:

P	TRUE(THAT(P)
A	A
R	R
D	D

This follows from the fact that the stronger member of the proposition that P is equivalent to the property of instantiating it, and the weaker member of the proposition that P is equivalent to the property of instantiating it. With these results in hand, all we have to do in order to derive the commitment table for the sentential truth predicate exhibited in section 1.4, is to define the sentential truth of S in the ordinary way – as the truth of the *x* such that S means *x*. (Actually, doing this requires that we have explicitly

[15] See chapter 11 of Schroeder [2008] for discussion.

introduced quantifiers into our expressivist object language, which I've omitted here for brevity's sake, but that is the basic idea.)

In other words, this shows how an expressivist framework can give a nondescriptivist account of truth which achieves the principal virtues catalogued when we began: it can explain full intersubstitutability, and indeed explain the 'pull' of every instance of T-Schema, without actually going so far as to endorse every such instance. It does this because it describes the commitment relationship codified by full intersubstitutability even though there is no conditional sentence which is capable of expressing this commitment relationship. And it makes good on the idea that the right response to liar-paradoxical sentences is to reject them – indeed, it is a theorem of the theory that anyone who accepts that Liar means that Liar is not true is committed to rejecting Liar. Moreover, no problem akin to Liar's Revenge arises, because the theory explains why it is not possible to lexicalize rejection, and so nothing like Liar's Revenge is expressible.

Finally, this theory provides a deflationary resolution to the paradoxes, by explaining why it is that there is nothing that we are missing out on, by rejecting Liar, rather than accepting or denying it. For on this theory, it is not propositions which play a role in carving up the world at its joints and consequently as the appropriate objects of excluded middle, but representational contents. For each representational content, the world must either be some way – such that that representation content obtains, or such that it does not. So if we reject a sentence like 'grass is green', which corresponds not only to proposition, but to a representational content, we miss out on something – something that we could have come to realize about the world, if we hadn't rejected. In contrast, no sentence involving 'true' corresponds directly to a representational content. Some sentences involving 'true' – the grounded ones – do have the feature that if we reject them, we will be committed to rejecting some ordinary descriptive sentence, and hence to missing out on some feature of the world. But other sentences involving 'true' – the ungrounded ones – are ones such that rejecting them does not make us miss out on anything.

2.4 revenge after all?

Now, I've made a big deal out of the fact that if rejection cannot be lexicalized, in the sense that there is no sentence of the language which expresses a state of rejection in the way that ordinary sentences express states of belief, then the problems associated with Liar's Revenge do not arise. Yet you might rightly be skeptical. Isn't it well known that every response to the liar paradox is subject to paradoxes of revenge? Shouldn't there be some way of doing something equivalent, even though we can't lexicalize rejection?

This thought is more sharply put in the following way: the theory that I've been discussing makes a recommendation about what to do to liar-paradoxical sentences. It says that you are rationally committed to rejecting them (assuming that you understand what they mean). So here is something that we obviously *can* lexicalize, insofar as we are able to formulate this theory: the feature of being such that anyone who understands you is rationally committed to rejecting you. Let's call this *rejectable*, for short. Shouldn't we then be able to do with 'rejectable' what we could do with 'not*'? The revenge sentence would look like this:

Attempted Revenge Attempted Revenge is rejectable or Attempted revenge is not true.

If you deny Attempted Revenge, then you are committed to accepting that Attempted Revenge is both not rejectable and not not true – i.e., that it is true. And hence, if you understand what Attempted Revenge means, you will be committed to accepting it. So you can't consistently deny it. So far, this sounds just like the ordinary paradox of revenge.

However, if you accept Attempted Revenge, and you understand what it means, then you will be committed to accepting that Attempted Revenge is true. But from Attempted Revenge and the assumption that Attempted Revenge is true, it follows that Attempted Revenge is rejectable. So if you accept Attempted Revenge, you will be committed to thinking that it is rejectable – i.e., that you are rationally committed to rejecting it. This is *not* the same as *being* committed to rejecting it. Both accepting and rejecting the same sentence involves the kind of self-inconsistency Gibbard calls disagreement. Accepting a sentence that you think you rationally ought to reject involves a different sort of incoherence – it is more like believing that you are irrational than like actually being irrational, or alternatively, more like believing you are inconsistent than like actually being inconsistent. Similarly, if you reject Attempted Revenge, you will be committed to rejecting that Attempted Revenge is rejectable. This is not inconsistent, either. It is like acting in a way without believing that acting in that way is rational.

The 'paradoxical' issues raised by Attempted Revenge are more like Moore's paradox than like outright inconsistency; accepting Attempted Revenge is like believing something and believing that you don't believe it, and rejecting Attempted Revenge is like believing something and not believing that you believe it. The kind of incoherence this exhibits is interesting and important, but it is different in kind from the kind of incoherence you get into if you either accept or deny the original Liar sentence.

Like other theories of truth, it is natural, when confronted with Attempted Revenge, to get into a hierarchy of responses. In addition to rejecting sentences which it would be inconsistent to accept or deny, perhaps the course of wisdom is also to reject sentences for which either accepting or denying them would lead you to accept that you are being irrational. Once we formulate that advice, of course, it will be possible to formulate new Attempted Revenge sentences such that accepting them involves believing that you believe yourself to be irrationally inconsistent. If that sounds bad (though of course it is not as bad as believing yourself to be rationally inconsistent or as actually being rationally inconsistent), new advice will be called for, to reject sentences which get you into such a predicament. This, I suppose, is a sort of hierarchy.

Still, I believe that this sort of hierarchy is importantly different from other sorts of hierarchy. The categories of being rationally inconsistent, believing oneself to be rationally inconsistent, believing oneself to believe oneself to be rationally inconsistent, and so on, are categories that we already *have* and which are of independent importance, rather than being an infinite succession of new semantic categories made up special-purpose for dealing with the liar.[16] If this is the worst kind of revenge that this treatment of the liar leads to, it is a kind of revenge that we can learn to live with – indeed, we had better learn to, because it is a problem that we have quite independently of the Liar.[17]

[16] One of the potential implications of Williamson's [2000] anti-luminosity argument, for example, is the importance of distinguishing among each of the categories in this hierarchy.

[17] Special thanks also to Barry Lam, Mike McGlone, Nikolaj Jang Pederson and Cory Wright, Matti Eklund and Alexis Burgess, Jake Ross, Andrew Alwood, Billy Dunaway, and Scott Soames, and to an audience at the University of Michigan in October 2008, especially to Allan Gibbard, Jason Stanley, and Anthony Gillies.

references

Avron, Arnon [1991]. ‘Natural 3-Valued Logics – Characterization and Proof Theory.’ *The Journal of Symbolic Logic* 56(1): 276-294.

Eklund, Matti [2002]. ‘Inconsistent Languages.’ *Philosophy and Phenomenological Research* 64: 251-275.

Field, Hatry [2008]. *Saving Truth from Paradox*. Oxford: Oxford University Press.

Gibbard, Allan [2003]. *Thinking How to Live*. Cambridge: Harvard University Press.

Gupta, Anil [2005]. ‘Do the Paradoxes Pose a Special Problem for Deflationism?’ In J.C. Beall and Bradley Armour-Garb, eds., *Deflationism and Paradox*. Oxford: Oxford University Press.

Richard, Mark [2008]. *When Truth Gives Out*. Oxford: Oxford University Press.

Schroeder, Mark [2008]. *Being For: Evaluating the Semantic Program of Expressivism*. Oxford: Oxford University Press.

Tappenden, Jamie [1993]. ‘The Liar and Sorites Paradoxes: Toward a Unified Treatment.’ *Journal of Philosophy* 90(11): 551-577.

Williamson, Timothy [2000]. *Knowledge and Its Limits*. Oxford: Oxford University Press.