

## expressivist truth

Expressivism and truth have had a rocky relationship; this paper is a move toward reconciliation. I'll show how to give a semantics for 'true' and 'false' in the most promising expressivist framework I know of<sup>1</sup>, and explain how the resulting marriage can benefit both parties. This is because expressivists need an account of truth, and expressivism about truth itself has certain attractions in its own right. In particular, I'll show in a rigorous way how expressivists can make good on the idea that valid arguments are truth-preserving, shed some light on the idea that truth is not a property, and explore an application to the paradox of the liar. But first, I need to explain what an expressivist semantics involves, and some of the background underlying the version of the view that I will depend on, here. The digression will take a while, but I promise a payoff at the end.

### I.1 expressivism as a nondescriptivist semantic framework

I prefer to understand expressivism as a development of a certain kind of *assertability-conditional* semantics. So instead of assigning sentences to *truth* conditions, an expressivist semantics assigns them to *assertability* conditions. The main idea is that just as it is the job of a syntactic theory to say when a sentence violates the *syntactic* rules of a language, it is the job of a semantic theory to say when a sentence violates the *semantic* rules of a language. But sentences don't necessarily violate the semantic rules of the language when they are false, and they don't necessarily comply with the semantic rules of the language when they are true. Take, for example, the case of someone who sincerely believes that John Kerry is president of the United States, because she stopped paying attention to electoral politics shortly before November 2004. When asked who is president, she says, 'John Kerry is president'. She is making some kind of mistake, but it isn't a *semantic* mistake: her mistake is rather about United States electoral politics. The reason she isn't making a semantic mistake is intuitively that she really does believe that John Kerry is president.

---

<sup>1</sup> See my *Being For: Evaluating the Semantic Program of Expressivism* for further details of and motivation for the expressivist semantics sketched here.

Reasoning like this leads to the idea that someone makes a *semantic* mistake just in case she says something that she doesn't really think. So if it is ultimate job of a semantic theory to characterize the conditions under which a sentence involves a semantic mistake, every sentence, 'P', should be associated with the condition that the speaker thinks that P.<sup>2</sup> That is the condition, after all, under which asserting the sentence is semantically okay. An assertability-conditional expressivist semantics takes this lesson to heart, and aspires to provide a semantics for a language by associating each sentence 'P' with what it is to think that P. The assertability condition for asserting 'P' is then that the speaker satisfies what it is to think that P.

Now, if for any value of 'P' whatsoever, what it is to think that 'P' is to stand in a single, uniform, relation – *belief* – to a truth-evaluable content – the proposition that P – then any old truth-conditional semanticist can assign sentences to these kinds of assertability conditions. To do so, she simply takes the truth-evaluable content to which her semantics assigns 'P', and says that the assertability condition for asserting 'P' is that the speaker stands in the belief relation to that truth-evaluable content. So expressivist semantics only gets *interesting*, if for some values of 'P', thinking that 'P' does not really consist, at bottom, in the same kind of underlying mental state.

So, for example, according to *noncognitivist expressivism*, in metaethics, thinking that stealing is wrong is importantly different from thinking that grass is green. It consists in bearing a certain negative attitude – call it *disapproval* – toward stealing. Similarly, expressivism about epistemic modals holds that thinking that Max might be in Minnesota is importantly different from thinking that Max is in Minnesota, by being an attitude of lower confidence toward the same content, rather than in being the same attitude toward a different content. Expressivism about indicative conditionals holds that thinking that if Clinton wins the Democratic nomination, then McCain will win the general election is importantly different from thinking that Clinton will win the Democratic nomination; a natural version of this view holds that the conditional thought consists in a high conditional credence in the consequent, rather than a high outright credence in a conditional content. And expressivism about truth is the view that thinking that it is true that grass is green is importantly different from thinking that grass is green. We'll see later on just what expressivists about truth might take this difference to be.

Let's say that thinking that grass is green is a matter of bearing a certain attitude toward what we might call a *representational content*, where we can understand representational contents as

---

<sup>2</sup> Ignoring, for now, the qualifications required where 'P' contains some context-dependent element.

corresponding to genuine ways of dividing up the world as it exists independently of our minds, and as the appropriate objects of the principle of excluded middle.<sup>3</sup> If we help ourselves to this terminological convention, then interesting expressivist views hold that these constructions – moral predicates, epistemic modals, conditionals, and the truth predicate – don't make any semantic contribution to representational contents. In our metaphysics, there is nothing about the furniture of the world which they help to carve up. Hence, an expressivist theory of truth could be characterized as making sense of the claim that truth is not a property.

## I.2 the embedding problem

Expressivism, it is well-known, faces a major problem in accounting for the semantics of complex sentences. It is all well and good to offer a semantic theory according to which the semantic assertability condition of 'stealing is wrong' is that the speaker disapproves of stealing. A successful assertability-conditional semantic theory needs to assign a semantic assertability condition to *every* sentence, not just to atomic ones. But an assertability-conditional framework in which the semantic assertability condition of 'P' is that the speaker thinks that P cannot simply apply the same sorts of compositional rules as an ordinary truth-conditional semantics would. A truth-conditional semantics assigns to any sentence, ' $\sim$ P', the *complement* of the truth-condition of 'P'. But if the assertability condition of 'P' is that the speaker thinks that P, the complement of those conditions is that the speaker does not think that P. But not thinking that P is not the same thing as thinking that  $\sim$ P – and an expressivist assertability-conditional semantics must assign to ' $\sim$ P' the condition that the speaker thinks that  $\sim$ P. So an expressivist semantics cannot simply apply to assertability conditions, the same compositional rules that other theorists have applied to truth-conditions. And if they can't do that, then what are they to do?

The answer, is that an expressivist semantics needs to be able to say, for an arbitrary sentence 'P', where we know what it is to think that P, what is involved in thinking that  $\sim$ P. And then the expressivist needs to appeal to properties of what it is to think that P and what it is to think that  $\sim$ P, in order to explain the semantic relationship between those two sentences. This is something that expressivists have had a very hard time in doing, even just for the special case of

---

<sup>3</sup> I'm being cautious here not to call these items 'propositions', because expressivists are going to distinguish representational contents from those things which are the objects of the attitudes and bearers of truth and falsity – the theoretical roles standardly assumed to be played by propositions. See sections 4.2 and 5.1 below, and chapter 10 of *Being For*.

negation. But an adequate expressivist semantics needs to do this for every compositional construction in natural languages.

One of the hardest parts of formulating such a constructive expressivist semantics, is that its rules have to apply equally well, whether they are applied to ordinary descriptive sentences like ‘grass is green’ or to the sentences for which the expressivist aims to provide a special account – such as moral sentences. In essence, that means that the mental states expressed by moral sentences have to be *similar enough* to the states expressed by ordinary descriptive sentences, that the same sorts of compositional rules can apply, and such that we can provide a uniform and general explanation of why a sentence and its negation are always inconsistent, no matter what kind of mental state that sentence expresses. It is essentially this reasoning, which I’ve explained in much more detail elsewhere, that leads to the semantic program of biforcated attitude semantics.<sup>4</sup>

## 2.1 biforcated attitudes

The main idea of biforcated attitude semantics, the version of expressivist semantics that I will work with, here, is that at *some* level, every sentence *does* express the same kind of mental state – states that I call *biforcated attitudes*, for reasons that I’ll explain in a moment. The notion of a biforcated attitude, however, is highly general, and allows both for a kind of biforcated attitude that is constituted by the relationship to representational contents that is involved in ordinary descriptive belief, as in the belief that grass is green, as well as for a kind of biforcated attitude that does not involve any relationship to a representational content – as is the case with what it is to think that stealing is wrong.

To understand what biforcated attitudes are, and how they can come in these different varieties, we start by introducing the attitude of *being for*. Being for is a very general positive attitude, which according to biforcated attitude semantics, is the fundamental building block of many interesting kinds of mental states. I take being for to be directly motivational, and so since I think we can treat things that agents *do* – actions, in some very broad sense – as properties of the agent, I will assume that being for takes properties for its objects, rather than propositions, and that when you are for some property, F, then other things being equal, you will be motivated to be F.

---

<sup>4</sup> For a much more rigorous presentation of this intuitive thought, see chapters 2-7 of *Being For*.

Beyond these assumptions, I make one more very important assumption about the attitude of being for. It is that being for shares with belief and intention, in contrast to states like supposing and wondering, the property of what I call being *inconsistency-transmitting*. Beliefs come into a special kind of rational conflict with one another – a kind of conflict that Allan Gibbard [2003] calls ‘disagreement’ – just in case their contents are inconsistent.<sup>5</sup> Intentions can come into a similar kind of rational conflict with one another, and do so, again, just in case their contents are inconsistent. There is not, however, any rational inconsistency in simultaneously supposing that P and that  $\sim P$ , or in simultaneously wondering whether P and whether  $\sim P$ . It is hard to characterize exactly what is involved in the special kind of rational conflict that Gibbard calls ‘disagreement’, and it is hard to understand exactly why belief and intention are subject to it, while supposing and wondering are not. I’m not going to address such questions, here. The assumption that being for is *inconsistency-transmitting* is the assumption that two states of being for disagree with one another, in Gibbard’s sense, just in case their contents are inconsistent.

A biforcated attitude is, descriptively enough, a state consisting in *two* states of being for, one of whose contents is strictly stronger than the other. So, for example, if you are, on the one hand, for both avoiding stealing and disapproving of stealing, and are also, on the other hand, for disapproving of stealing, then you are in two states of being for, one of which is strictly stronger than the other. So you are in a biforcated attitude, a state which I will write like:

$$\langle \text{FOR}(\lambda z(\text{avoiding}(z, \text{stealing}) \wedge \text{disapproving}(z, \text{stealing})))^*, \text{FOR}(\lambda z(\text{disapproving}(z, \text{stealing}))) \rangle$$

This notational convention employs small-caps (‘FOR’) to denote mental states, lambda-abstractions on the dedicated variable z ( $\lambda z(\dots)$ ) to denote properties, and an asterisk, by convention, to designate the state of being for whose content is strictly stronger. This state of being for I’ll call the *major attitude* of the biforcated attitude, and the other, I’ll call the *minor attitude*.<sup>6</sup>

---

<sup>5</sup> This is not to say that it is always irrational to have inconsistent beliefs. Irrationality is a summary concept, to which many factors may contribute. It is just to say that there is a special kind of *clash* between beliefs with inconsistent contents and intentions with inconsistent contents that does not occur between supposings with inconsistent contents or wonderings with inconsistent contents. Gibbard associates this special kind of clash with the interpersonal test given by our intuitive notion of disagreement, in the sense that if you have a belief and I have a belief with an inconsistent content, then we *disagree*. But this interpersonal test may or may not perfectly line up with the property that I am after; for my purposes Gibbard’s intrapersonal test, which is given by the question of whether switching from one state to the other counts as a *change of mind*, in an intuitive sense, is more appropriate.

<sup>6</sup> In *Being For*, I didn’t require that the major property be *strictly* stronger than the minor property, only that it be at least as strong. This led to a problem that I there called the *new new negation problem*. The present requirement avoids that problem, and is a necessary assumption for one of my key claims about the liar paradox in section \*\*\*\*.

One natural way to develop noncognitivist metaethical expressivism, within the framework of biforcated attitude semantics, is to identify the biforcated attitude that we have just discussed, with what it is to think that stealing is wrong. On this view, to think that stealing is wrong is to be for disapproving of stealing and avoiding it, and for disapproving of stealing. So this view predicts that someone who thinks that stealing is wrong will be motivated, other things being equal, to disapprove of stealing and avoid it. Intuitively, that is not very far off of what we might expect someone who thinks that stealing is wrong to be motivated to do. So this is a picture of what it is to think that stealing is wrong that can explain its essential connection both to action, and to the emotions.

## 2.2 descriptive sentences and disbelief

I have just shown how an intuitive desire-like attitude, such as that which metaethical noncognitivists have held to be involved in thinking that stealing is wrong, can be understood as a biforcated attitude. But we can also understand ordinary descriptive belief, as is involved in thinking that grass is green, as a biforcated attitude. To see how, start by asking what we expect someone who thinks that grass is green to be motivated to do, other things being equal. Intuitively, we would expect such a person to take grass's being green as settled, in deciding what to do. I'm going to abbreviate this relation by calling it *proceeding as if*, and identify thinking that grass is green, with being for proceeding as if grass is green, and also being for not proceeding as if grass is not green, where I understand the objects of proceeding as if to be what I was earlier calling representational contents. On the assumption that it is impossible to both proceed as if some representational content and also proceed as if its negation, but that it is possible to not proceed as if either,<sup>7</sup> this is also a biforcated attitude, and I'll write it like this:

$$\langle \text{FOR}(\lambda z(\text{pai}(z, \text{that}(\text{green}(\text{grass}))))), \text{FOR}(\lambda z(\neg \text{pai}(z, \text{that}(\neg \text{green}(\text{grass})))))) \rangle$$

Again, to get a flavor for how the notation works, I'm abbreviating 'proceeding as if' by 'pai', and using 'that' as a term-forming operator which designates the representational content that we would ordinarily associate with its argument. I'm assuming that the language that is used to state these biforcated attitudes all comes from a purely descriptive metalanguage each of whose sentences

---

<sup>7</sup> The two parts of this assumption are what are required in order to ensure the result that the property of proceeding as if  $p$  is strictly stronger than the property of not proceeding as if  $\neg p$ .

can be associated with what I'm calling a representational content, and I'm reserving 'pointy' connectives ( $\neg$ ,  $\wedge$ ,  $\forall$ ) for this metalanguage; in a little bit I'll state an expressivist semantics for a simple object-language, for which I'll use 'rounded' connectives ( $\sim$ ,  $\&$ ,  $()$ ).

It is an important, and I think attractive, feature of biforcated attitude semantics, that it allows us to make sense of a distinct attitude of *disbelief* or *doubt*, which is rationally inconsistent with belief, but still weaker than belief in the negation. This is an important feature that will come in later. To doubt that grass is green is not to believe that grass is not green, but it does rationally conflict with believing that grass is green. If believing that grass is green is being for proceeding as if grass is green and for not proceeding as if grass is not green, then believing that grass is not green is being for proceeding as if grass is not green and for not proceeding as if grass is not not green. This is a biforcated attitude which has two parts, a major attitude, and a minor attitude. Its minor attitude is the state of being for not proceeding as if grass is not not green – i.e., as if it is green. Someone can be in this state without believing that grass is not green, because she may not be in the major attitude of believing that grass is green. But it is still rationally inconsistent with believing that grass is green, because it is a state of being for which 'disagrees', in Gibbard's sense, with the major attitude of believing that grass is green. (This is because both are states of being for, and their contents are inconsistent, and being for is inconsistency-transmitting.)

I'll come back to explain more about disbelief in a little bit, and to generalize. But first, we need to show how the theory of biforcated attitudes can be used in order to state a constructive, compositional expressivist semantics, which allows for a formally adequate explanatory account of why complex sentences have the appropriate semantic properties.

### 3.1 biforcated attitude semantics

We already have the main pieces of notation that we need in order to state our biforcated attitude semantics for a simple language with both moral and ordinary descriptive predicates. The only new piece of terminology in what follows, is that I introduce the notion of a *semantic value*. This is a technical notion, whose theoretical role is exhausted by the stipulation that only closed sentences express states of mind, and for any closed sentence 'A' with semantic value  $\langle \lambda z(\alpha^1)^*, \lambda z(\alpha^2) \rangle$ , 'A' expresses the biforcated attitude,  $\langle \text{FOR}(\lambda z(\alpha^1)^*), \text{FOR}(\lambda z(\alpha^2)) \rangle$ . So intuitively, we use semantic values to construct the pairs of properties that turn out to be the contents of the two states of

being for that are involved in the biforcated attitude expressed by a sentence. We start with the semantics for a few predicates:

<b>WRONG</b>	‘WRONG(x)’ is a well-formed formula and has the semantic value $\langle \lambda z(\text{avoiding}(z,x) \wedge \text{disapproving}(z,x))^*, \lambda z(\text{disapproving}(z,x)) \rangle$ .
<b>BETTER</b>	‘BETTER(x,y)’ is a well-formed formula and has the semantic value $\langle \lambda z(\text{choosing}(z,x,y) \wedge \text{preferring}(z,x,y))^*, \lambda z(\text{preferring}(z,x,y)) \rangle$ .
<b>COMMON</b>	‘COMMON(x)’ is a well-formed formula and has the semantic value $\langle \lambda z(\text{pai}(z, \text{that}(\text{common}(x))))^*, \lambda z(\neg \text{pai}(z, \text{that}(\neg \text{common}(x)))) \rangle$
<b>RARER</b>	‘RARER(x,y)’ is a well-formed formula and has the semantic value $\langle \lambda z(\text{pai}(z, \text{that}(\text{rarer}(x,y))))^*, \lambda z(\neg \text{pai}(z, \text{that}(\neg \text{rarer}(x,y)))) \rangle$

(Notice that ‘COMMON’ and ‘RARER’ sentences are going to express states of mind that are constituted by a relation to representational contents, but that ‘WRONG’ and ‘BETTER’ sentences will express states that are constituted by no such relation. This corresponds to the idea that the former are ‘descriptive’ predicates, and the latter ‘nondescriptive’.) We can then state the following compositional rules:<sup>8</sup>

~	If ‘A’ is a well-formed formula with semantic value $\langle \lambda z(\alpha^1)^*, \lambda z(\alpha^2) \rangle$ , then ‘~A’ is a well-formed formula and has semantic value $\langle \lambda z(\neg \alpha^2)^*, \lambda z(\neg \alpha^1) \rangle$ .
&	If ‘A’ is a well-formed formula with semantic value $\langle \lambda z(\alpha^1)^*, \lambda z(\alpha^2) \rangle$ , and ‘B’ is a well-formed formula with semantic value $\langle \lambda z(\beta^1)^*, \lambda z(\beta^2) \rangle$ , then ‘A&B’ is a well-formed formula and has semantic value $\langle \lambda z(\alpha^1 \wedge \beta^1)^*, \lambda z(\alpha^2 \wedge \beta^2) \rangle$ .
()	If ‘A’ is a well-formed formula with semantic value $\langle \lambda z(\alpha^1)^*, \lambda z(\alpha^2) \rangle$ , then ‘(x)(A)’ is a well-formed formula and has semantic value $\langle \lambda z(\forall x(\alpha^1))^*, \lambda z(\forall x(\alpha^2)) \rangle$ .
names	If ‘a’ is a referring term whose referent is o and ‘AxB’ is a well-formed formula open in x whose semantic value is $\langle \lambda z(\alpha_x \beta) \rangle^*$ ,

<sup>8</sup> In what follows I treat ‘ $\alpha$ ’ and ‘ $\beta$ ’ as schematic letters standing in for arbitrary metalanguage formulas and ‘A’ and ‘B’ as schematic letters standing for arbitrary object-language formulas.

$\lambda z(\alpha x \beta)$ ), then ‘AaB’ is a well-formed formula and has semantic value  $\langle \lambda z(\alpha o \beta)^*, \lambda z(\alpha o \beta) \rangle$ .

Finally, let’s specify that ‘murder’ and ‘stealing’ are referring terms in this simple language, with the obvious referents. Intuitively, what each compositional rule tells us to do, is to ‘push the connective inside’ the state of being for – pairing major attitudes with major attitudes, and minor with minor, in the case of ‘&’. It is an important fact – verifiable by a simple induction on formula complexity – that this semantics allows only for sentences that express biforcated attitudes – pairs of states of being for, one of which is strictly stronger than the other.

This simple semantics is both constructive and compositional. Existing expressivist views typically work by saying that complex sentences express that state of mind – whatever it is – which has the right inferential relationships between other states of mind, for us to use those inferential relationships in order to predict that it has the right semantic properties. But they don’t actually tell us what state of mind that is, nor explain why it has the right inferential relationships to other states of mind. Biforcated attitude semantics, in sharp contrast, tells us exactly what biforcated attitude is expressed by each and every complex sentence in the language, by telling us how to compose the contents of its major and minor attitudes from the contents of the major and minor attitudes of its parts. We can then use the single, very general, assumption that being for is inconsistency-transmitting, in order to *explain* the inferential relationships among arbitrary sentences of this language, in a way that I’ll explain more of, in the next section.<sup>9</sup>

### 3.2 logic in biforcated attitude semantics

Given this semantics, we can now say that to *accept* ‘P’ is to be in the biforcated attitude expressed by ‘P’, to *deny* ‘P’ is to be in the biforcated attitude expressed by ‘ $\sim$ P’, and that to *disaccept* ‘P’ – the generalization of disbelief – is to be in the minor attitude of ‘ $\sim$ P’. We can then define *disacceptance full-stop* of ‘P’ as disaccepting both ‘P’ and ‘ $\sim$ P’. (It is important to note that disacceptance full-stop is not, in general, itself a biforcated attitude.) For any sentence ‘P’, each of these attitudes is coherent and rationally conflicts with each of the others – so they form a partition of the possible fully opinionated responses that it is possible to have to ‘P’. We can also characterize these states of mind by their *for-commitment classes*. If ‘P’ expresses the biforcated attitude whose contents are the pair,  $\langle \pi^1, \pi^2 \rangle$ , we can say that to accept ‘P’ is to be for-committed to  $\{\pi^1, \pi^2\}$ , that to deny ‘P’ is to

---

<sup>9</sup> For further details and discussion, see chapters 8-10 of *Being For*.

be for-committed to  $\{\neg\pi^1, \neg\pi^2\}$ , that to disaccept ‘P’ full-stop is to be for-committed to  $\{\neg\pi^1, \pi^2\}$ , that to *merely disaccept* ‘P’ is to be for-committed to  $\{\neg\pi^1\}$ , that to *merely disdeny* ‘P’ is to be for-committed to  $\{\pi^2\}$ , and that to *withhold* ‘P’ is to be for-committed to  $\{\}$ .

Intuitively, disacceptance full-stop is an attitude that it makes sense to have toward paradoxical sentences. If it is paradoxical, then don’t accept it, and don’t deny it. Just disaccept it full-stop. Once I introduce a semantics for ‘true’ and am able to construct some paradoxical sentences in a language governed by biforcated attitude semantics, we’ll be able to make this idea more precise, to illustrate how it works, and to prove that in fact, disacceptance full-stop is always both a coherent and the only coherent attitude to have toward liar-sentences and their progeny.

It is a theorem of biforcated attitude semantics, that all and only the sentences which are theorems of classical logic are inconsistent to deny under any uniform interpretation of their predicates and referring terms within biforcated attitude semantics. That is, for any classical theorem, its negation expresses a biforcated attitude whose major attitude disagrees, in Gibbard’s sense, with its minor attitude. Since major and minor attitudes are always states of being for, and states of being for disagree just in case their contents are inconsistent, this is something that we can actually prove, given the assumptions that we’ve made so far. I’ll omit the proof here.<sup>10</sup>

So the theorems of classical logic are sentences that are in this sense undeniable: that denying them involves the same kind of rational inconsistency as believing or intending inconsistent things. That sounds good – this is an important feature of ‘logical truths’. For our purposes later, we can also generalize, and introduce the notion of an S-theorem, where S is a set of words in our object language. Let an S-theorem be a sentence that is inconsistent to deny under any biforcated attitude semantics for every word which appears in that sentence, other than the words which are in S. So if L is the set,  $\{\sim, \&, ()\}$ , then our result so far is that all and only the theorems of classical logic are L-theorems. Intuitively, any sentence that is an  $L \cup \{F\}$ -theorem, for some word, ‘F’, but not an L-theorem, is one whose status as undeniable is due to the meaning of ‘F’. This is a concept that will come in handy in just a little bit, and it is natural to identify such sentences with non-logical analyticities, in some broad sense.

Classically valid arguments have the feature that the conjunction of their premises with the negation of their conclusion is a logical falsehood – i.e., its negation is a logical truth. So it follows from our theorem that all and only classically valid arguments are ones such that it is inconsistent

---

<sup>10</sup> See chapter 8 of *Being For*.

to accept their premises and deny their conclusions. It does not turn out in biforcated attitude semantics, however, as one might expect, that accepting the premises of a valid argument necessarily *commits* you to accepting its conclusion, in the sense that every state of mind that disagrees with the state of mind expressed by the conclusion already disagrees with the state of mind expressed by the conjunction of the premises. For example, the inference from an arbitrary or null premise to any logical truth is classically valid. But though logical truths are undeniable, it does not follow in biforcated attitude semantics that they must be accepted. If you disaccept ‘P’ full-stop, for example, then you are committed to disaccepting ‘ $\sim(P \& \sim P)$ ’ full-stop. So only relevance-valid arguments commit someone who accepts their premises to accepting their conclusion – and that is an attractive result.<sup>11</sup>

You might think, however, that this account leaves something important unexplained about logically valid arguments. It is one thing, after all, to state a criterion that is formally adequate, in the sense that it captures all and only the right arguments as ‘valid’. And it is one thing for this criterion to intuitively have *something* to do with validity, because it captures our sense that there is a kind of rational inconsistency in accepting the premises of a valid argument and denying its conclusion. But it is another thing to explain why an argument is *valid*, you might think. After all, valid arguments are *truth-preserving*.

Since at least Simon Blackburn, expressivists have been interested in the idea that they can ‘earn the right to truth’. What this means, is that they have aspired to give an explanatory and formally adequate account of some *other* feature of valid arguments, such as the one we have picked out in this section, and then to *use* that account, along with a semantics for the truth-predicate, in order to *derive* the result that valid arguments are truth-preserving. This is a promise that I will show how to make good on in this paper, but first we need to introduce an appropriate semantics for ‘true’.

#### 4.1 minimalism about truth

The basic idea behind minimalist approaches to truth, to which expressivists generally seek to turn at this point, is roughly that the meaning of ‘true’ is more-or-less exhausted by the appropriateness of instances of ‘Schema T’: ‘It is true that P just in case P.’ The problem with this, however, is that not all instances of Schema T *are* appropriate. To begin with, the paradox of the liar and its

---

<sup>11</sup> See chapter 8 of *Being For* for further discussion.

strengthened versions show that Schema T must be restricted in *scope* – a restriction that it is painfully complicated to spell out precisely, and which it is hard to fathom is really part of our grasp of truth as competent speakers.

Things are even more complicated for expressivists who seek to take advantage of a minimalist account of truth; as I just formulated it, Schema T is a schema for ‘propositional truth’ – it applies to ‘that’ clauses. But expressivists have not had any straightforward way of understanding the semantic role of ‘that’ clauses, which makes it hard for them to interpret what Schema T says, let alone to generalize on it. This has generally meant that expressivists have been inclined to stick to accounts of sentential truth, formulating Schema T as the claim that ‘S’ is true just in case S. But this, again, requires further restrictions. If ‘S’ itself has any context-dependent elements, then this only tells us when ‘S’ is true relative to our *own* context of utterance – it tells us nothing about when ‘S’ is true relative to alternative contexts of utterance. But since it tells us nothing about when ‘S’ is true relative to alternative contexts of utterance, it doesn’t suffice for a minimalist account of truth – on the contrary, we need to add something else to the theory in order to know how ‘true’ applies to sentences in other contexts. Alternatively, we could restrict the *scope* of the Schema T generalization again, to context-independent sentences, but again, that leaves us something short of a minimalist account of truth – we need to *add* something, in order to know how to assess context-dependent sentences for truth.

Restricting our account of truth to sentential truth also still leaves us with the important problem of extending that account to ordinary ‘propositional’ uses, which would ordinarily be treated by treating truth as predicated of a proposition. Difficulties with not taking this surface structure seriously derive from the entailment relations among ‘Max said that P’, ‘P is true’, and ‘something Max said is true’. These entailment relations are the main reason why philosophers so often treat ‘that’ clauses as denoting propositions, which are both the bearers of truth and falsity, and the objects of attitudes like belief and assertion.

In addition to these constraints, it is important that ‘it is true that P’ should turn out to mean something different from ‘P’, not only because creatures without a concept of truth may think the latter without thinking the former, but because the predicate ‘true’ in ‘it is true that P’ has greater generality, and can be applied even in ignorance of what sentence would express the thought that is being said to be true. For example, as in ‘what he said is true’ or in ‘everything written in that document is true’.

Straightforward redundancy accounts of ‘true’ run into difficulties with these features of ‘true’, which is part of why minimalists generally depart from the idea that ‘it is true that P’ means the same thing as ‘P’. Instead, they offer us the idea that ‘true’ is a ‘device of generalization’ on asserting ‘P’ (which is a way of stipulating that it is like a redundancy account, except that it doesn’t have these problems). Or they say that it comes with a ‘rule for use’, to be used in a situation in which you would assert ‘P’. Or they deny that truth is a property, but don’t quite tell us what it is, instead. These ideas, though attractive, are vague and schematic, and it is hard to test their consequences in a rigorous way. In some ways, they are strikingly like the proposals in many existing expressivist theories that complex sentences express those states of mind, whatever they are, which have the right inferential properties, without telling us what states of mind those are, or explaining why they have those inferential properties.

If minimalist ideas about the semantics for ‘true’ are sometimes vague and difficult to spell out fully or test, expressivist appeals to minimalism have been even more tenuous. Simon Blackburn and others have gestured to the effect that expressivists can ‘earn the right’ to truth by appeal to some sort of minimalist account of ‘true’. But they have not generally offered a concrete minimalist semantics for ‘true’ or shown how it yields appropriate and successful semantic predictions. In what follows I show that within the semantic framework of biforcated attitude semantics, it is possible to give an elegant, quite simple, account of the semantics of ‘true’ and ‘false’ that is minimalist without being redundant, that makes sense of ‘propositional’ truth, along with the corresponding sorts of entailments, that puts flesh on the idea that truth is not a property, and that is rigorous enough to yield, in connection with the rest of the framework of biforcated attitude semantics, concrete and testable predictions – predictions which are to the benefit of both parties.

#### **4.2 bringing biforcated attitude semantics to bear**

The first step in developing an expressivist semantics for sentences like ‘it is true that P’ is to ask what semantic role ‘that P’ plays. Because ‘it is true that P’ and ‘Max said that P’ together imply ‘Max said something true – namely, that P’, it is natural to treat ‘that P’ as a referring term. Expressivists have had a hard time sorting out what ‘that P’ is to refer to, but the semantic framework of biforcated attitude semantics spits out a natural answer. Because of the extra structure created by the move that allowed us to build a constructive and compositional semantics,

biforcated attitude semantics allows us to distinguish what I've been calling the *semantic value* of a sentence from the mental state that it *expresses*. In biforcated attitude semantics, each closed sentence has as its semantic value a pair of properties, one of which is strictly stronger than the other. So in general, where 'P' is a sentence, we can treat 'THAT(P)' as a singular term denoting the semantic value of 'P':

**THAT**            If 'P' is a well-formed formula, then 'THAT(P)' is a referring term. Relative to each assignment of values to the unbound variables in 'P', it refers to the semantic value of 'P' relative to that assignment.<sup>12</sup>

The virtue of this approach is not only that it will allow us to make sense of the aforementioned entailments in the ordinary way, but that it will allow our minimalist account to generalize immediately to sentences like 'what Max said is true' and 'everything Max believes is true' – and not merely by stipulation that we intend it so to generalize. In fact, to warm up for our expressivist treatment of 'true' and 'false', we can first deal with the somewhat simpler case of providing an expressivist semantics for 'thinks that'. The main idea is simple. Since 'THAT(P)' denotes the semantic value of 'P', the state expressed by 'P' is the biforcated attitude of being for each of the properties in the semantic value of 'P', and in general, the state expressed by 'P' is what it is to think that P, it follows that 'THINKS(Max,THAT(P))' should be an ordinary descriptive sentence which says, intuitively, that Max is for each of the properties in the semantic value of 'P'. Given that ordinary descriptive sentences express beliefs, and generalizing on the model provided by 'common' and 'rarer', from section 3.2, we can encapsulate this as follows ('jointfor(a,b)' says that a is for each of the properties in b, where b is a pair of properties, one of which is strictly stronger than the other):

**THINKS**            'THINKS(x,y)' is a well-formed formula and has the semantic value  $\langle \lambda z(\text{pai}(z, \text{that}(\text{jointfor}(x,y)))) * \lambda z(\neg \text{pai}(z, \text{that}(\neg \text{jointfor}(x,y)))) \rangle$

Together with our clause for 'THAT(P)', above, this provides a semantics for sentences like 'THINKS(Max,THAT(P))', intuitively glossable as 'Max thinks that P'. But it also generalizes, and

---

<sup>12</sup> I'm being somewhat sloppy here to avoid digressions; strictly speaking I haven't given open sentences semantic values relative to assignments, I instead simply had a closure rule that allowed us to replace variables with names. This account is modified slightly from the version that I state in chapter 10 of *Being For* to allow for quantification into attitude and truth ascriptions.

allows for sentences like ‘Max thinks everything that Caroline thinks:  $\forall x(\sim(\text{THINKS}(\text{Caroline}(x))\&\sim(\text{THINKS}(\text{Max},x))))$ ’.

The approach we are exploring here, which treats ‘THAT(P)’ as a complex term referring to the semantic value of ‘P’, and treats ‘thinks’ and ‘true’ as predicates applying to semantic values, can be thought of as advocating a divorce among the theoretical roles commonly thought to be played by propositions – that they are not only the objects of the attitudes and bearers of truth and falsity, but that they play a role in dividing up the world at its joints – and hence in metaphysical commitment – and are the appropriate objects of excluded middle. The biforcated attitude semanticist will say, once she has adopted this semantics for ‘thinks’ and for ‘true’, that there are things which agents think and which can be true or false – and hence which are the objects of the attitudes and bearers of truth and falsity. These things are not, however, representational contents, and do not carry any metaphysical commitment. They are just pairs of properties – the kinds of thing assigned to each sentence by our compositional semantics. (They are simply the things that someone who thinks that P is thereby disposed to *do*, other things being equal.) Hence, they need to be distinguished from representational contents, which are the objects of some, but not all, kinds of thought.

On this view, ‘that grass is green’ does not denote a representational content, but in thinking that grass is green, someone does bear a special relation to a representational content. It is simply not the case that every kind of thought involves such a relation. So for every sentence, ‘P’, there is a proposition that P – a referent of ‘that P’ – and we need no ‘deflationary’ reading of the existential quantifier in order to make this claim. But for only some sentences, ‘P’, is there a corresponding representational content.

## 5.1 propositional truth

The account of ‘true’ builds on these basic ideas, but avoids committing to the idea that ‘true’ sentences express ordinary descriptive beliefs. Instead, for an arbitrary sentence ‘P’ with semantic value  $\langle \pi^1, \pi^2 \rangle$ , what we need to look for, is some pair of properties that are *equivalent* to but not *identical* to  $\pi^1$  and  $\pi^2$ . The most natural choices are the properties of *instantiating*  $\pi^1$  and *instantiating*  $\pi^2$ . This is the pair of properties that we want to make the semantic value of ‘P’. In general, let ‘ $T^{\text{major}}(a,b)$ ’ say that b is a pair of properties one of which is strictly stronger than the other and that a instantiates the stronger member of this pair, let ‘ $T^{\text{minor}}(a,b)$ ’ say that b is a pair of properties one

of which is strictly stronger than the other and that a instantiates the weaker member of this pair, and let ‘neg(x)’ denote the pair of properties whose members are the negations of the members of the pair of properties, x. With that notation, we can now state the expressivist accounts of ‘true’ and ‘false’.

**TRUE** ‘TRUE(x)’ is a well-formed formula and has semantic value  $\langle \lambda z(I^{\text{major}}(z,x)), \lambda z(I^{\text{minor}}(z,x)) \rangle$ .

**FALSE** ‘FALSE(x)’ is a well-formed formula and has semantic value  $\langle \lambda z(I^{\text{major}}(z,\text{neg}(x))), \lambda z(I^{\text{minor}}(z,\text{neg}(x))) \rangle$ .

Together with the account of ‘THAT(P)’, above, this account predicts that if ‘P’ is a closed sentence with semantic value  $\langle \pi^1, \pi^2 \rangle$ , then ‘TRUE(THAT(P))’ is a well-formed formula and has semantic value  $\langle \lambda z(I^{\text{major}}(z, \langle \pi^1, \pi^2 \rangle)), \lambda z(I^{\text{minor}}(z, \langle \pi^1, \pi^2 \rangle)) \rangle$ .

Now if we define ‘ $\supset$ ’ and ‘ $\equiv$ ’ in the standard way so that ‘ $P \supset Q$ ’ is an abbreviation for ‘ $\sim(P \& \sim Q)$ ’ and ‘ $P \equiv Q$ ’ is an abbreviation for ‘ $(P \supset Q) \& (Q \supset P)$ ’, ‘TRUE(THAT(P)) $\equiv$ P’ turns out to be an  $L \cup \{ \text{‘TRUE’, ‘THAT’} \}$ -theorem. (See appendix A.I.) Intuitively, logic plus the meaning of ‘TRUE’ and ‘THAT’ guarantee that ‘TRUE(THAT(P)) $\equiv$ P’ is undeniable, for any value of ‘P’. So given the meaning of ‘TRUE’ and ‘THAT’, ‘TRUE(THAT(P))’ is guaranteed to be equivalent to ‘P’. But this account is also not redundant - in particular, someone can think that  $p$  is true without thinking that  $p$ .

The account also generalizes immediately to allow uses of ‘true’ in sentences that refer only indirectly to the object of the truth-ascription – for example, as in ‘everything Max thinks is true’: ‘(x)(THINKS(Max,x) $\supset$ TRUE(x))’. Moreover, in contrast to some approaches to minimalism, it does so not simply by stipulating that ‘true’ ‘works like a predicate’, but by actually spelling out the underlying semantics, from which it *follows* that ‘true’ works like a predicate. It also predicts the right entailment relations among the sorts of sentences involving ‘that’ phrases. For example, from ‘Max thinks everything Caroline thinks’ and ‘everything Max thinks is true’, follows ‘everything Caroline thinks is true’ – for the very same reasons that any argument of the form ‘(x)(F(x) $\supset$ G(x))’, ‘(x)(G(x) $\supset$ H(x))’, ‘(x)(F(x) $\supset$ H(x))’ is valid – a feature that biforcated attitude semantics independently allows us to explain, along the lines indicated earlier. All of these are *very nice things*.

This account of truth also makes very nice sense of the idea that truth is *not a property*. Truth is like wrongness in not being a property, because like ‘wrong’ and ‘better’, our expressivist semantics treats ‘true’ in an *essentially expressivist* fashion, by treating atomic sentences formed using these predicates as expressing states that do not consist in the same sort of relation to representational contents as ordinary descriptive beliefs, such as the belief that grass is green. Ordinary descriptive sentences, in biforcated attitude semantics, express states of being for proceeding as if  $p$ , and being for not proceeding as if  $\neg p$ , where  $p$  is some representational content. So a predicate, ‘F’ can be thought of as denoting *properties*, in some strict sense, just in case the semantic value of ‘F(x)’ consists in the pair,  $\langle \lambda z(\text{pai}(z, \text{that}(f(x))))^*, \lambda z(\neg \text{pai}(z, \text{that}(\neg f(x)))) \rangle$ , for some suitable value of ‘f’. Intuitively, in such a case,  $f$  is the property picked out by ‘F’. But ‘true’ doesn’t work like that. Since the semantics treats it differently from ordinary descriptive predicates like ‘common’ and ‘rarer’, that puts meat on the idea that truth is not a property, and explains why there *need be* no property of truth, in order for the semantics to do everything that it needs to do, and in a rigorous fashion, which makes concrete and testable predictions.

## 5.2 earning the right to truth

The foregoing expressivist account of ‘true’ allows us to show that, *given* the account of validity from section 3.2, all and only valid arguments are provably truth-preserving. That is, we can show that for any valid argument, it is a theorem that if its premises are each true, then its conclusion is true. We show that by showing that if an argument with premises ‘P<sub>1</sub>’-‘P<sub>n</sub>’ and conclusion ‘C’ is valid, then  $(\text{TRUE}(\text{THAT}(P_1)) \& \dots \& \text{TRUE}(\text{THAT}(P_n))) \supset \text{TRUE}(\text{THAT}(C))$  is an  $\text{L}\cup\{\text{‘TRUE’}, \text{‘THAT’}\}$ -theorem – that is, that its undeniability is guaranteed by the meaning of ‘TRUE’ and ‘THAT’. On the assumption that biforcated attitude semantics is really correct about English, therefore, this is a kind of transcendental argument that if the premises of a valid argument are true, then its conclusion is, as well. This is because it shows us that the sentence which says so has the same status as sentences which state logical truths.

Here is the proof: take any valid argument, with premises ‘P<sub>1</sub>’-‘P<sub>n</sub>’ and conclusion ‘C’. It is valid just in case the conjunction, ‘P<sub>1</sub> & ... & P<sub>n</sub> & ~C’, is inconsistent to accept. But since ‘P’ and ‘TRUE(THAT(P))’ are equivalent, this conjunction is inconsistent to accept just in case the conjunction, ‘TRUE(THAT(P<sub>1</sub>)) & ... & TRUE(THAT(P<sub>n</sub>)) & TRUE(THAT(~C))’ is inconsistent to accept. But ‘TRUE(THAT(~C))’ is also equivalent to ‘~TRUE(THAT(C))’, on the semantics that I’ve given, so

the former conjunction is inconsistent to accept just in case ‘TRUE(THAT(P<sub>1</sub>))&...&TRUE(THAT(P<sub>n</sub>))&~TRUE(THAT(C))’ is inconsistent to accept – that is, just in case ‘~((TRUE(THAT(P<sub>1</sub>))&...&TRUE(THAT(P<sub>n</sub>)))&TRUE(THAT(~C)))’ is a theorem.<sup>13</sup> So all and only valid arguments are provably truth-preserving. In this way, an account like this one can make good on Blackburn’s promise, never carried out in detail, to ‘earn the right’ to truth. Though we began by giving an alternative characterization of validity, we still end up being able to explain and capture the idea that valid arguments are the truth-preserving ones. Truth really does have something to offer expressivism.

## 6.1 meaning and sentential truth

I think that expressivism also has something to offer truth, over and above the basic idea so far, which involved showing how to give a constructive semantics for ‘true’ which, rather than *stipulating* the inferential properties of sentences involving ‘true’, gave a concrete characterization of the mental state that is involved in accepting any such sentence, and predicted the inferential properties of those sentences on the basis of more general assumptions in the philosophy of mind – namely, that being for is inconsistency-transmitting. Given these features, biforcated attitude semantics constitutes a robustly explanatory nondescriptivist semantics for ‘true’, which not only makes concrete predictions, but explains them by appeal to a (relatively) small set of assumptions in the philosophy of mind. But beyond these nice features, I think that biforcated attitude semantics can also shed some light on some of the puzzles about truth. To see how, I’ll first introduce a sentential truth predicate.

Intuitively, we want it to turn out that if S means that P, then S is true just in case it is true that P – i.e., just in case P. So before I introduce the sentential truth predicate, I first want to introduce the natural semantics for ‘means that’. Because ‘means that’ is intuitively an ordinary descriptive relation – the relation a sentence bears to its semantic value – this leads immediately to the following rule, where ‘sv(x)’ designates the semantic value of x:

**MEANS**            ‘MEANS(x,y)’ is a well-formed formula and has the semantic value  
 $\langle \lambda z(\text{pai}(z, \text{that}(sv(x)=y)))^*, \lambda z(\neg \text{pai}(z, \text{that}(\neg sv(x)=y)))) \rangle$

---

<sup>13</sup> This argument, is slightly different from the argument that I give in chapter II of *Being For*, which only established one direction of this equivalence.

So in particular, ‘MEANS(S,THAT(P))’ expresses the ordinary descriptive belief with the representational content that S’s semantic value is  $\langle \pi^1, \pi^2 \rangle$ , where  $\langle \pi^1, \pi^2 \rangle$  is the semantic value of ‘P’.

Now we can state the rule for our sentential truth-predicate (‘the x:(...)’ means precisely what it looks like):

**true**      ‘true(x)’ is a well-formed formula and has the semantic value  
 $\langle \lambda z(I^{\text{major}}(z, \text{the } y:(\text{pai}(\text{sv}(x)=y))))^*, \lambda z(I^{\text{minor}}(z, \text{the } y:(\neg \text{pai}(\text{that}(\neg \text{sv}(x)=y)))))) \rangle$

This definition makes it easy to prove that every instance of the following schema (where ‘s’ is replaced by a name for a sentence and ‘P’ is replaced by a sentence) is an  $\text{LU}\{\text{‘true’}, \text{‘MEANS’}, \text{‘THAT’}\}$ -theorem (see Appendix A.2 for the proof):

**T-schema**       $\text{MEANS}(s, \text{THAT}(P)) \supset (\text{true}(s) \equiv P)$

What this means, is that given the meanings of ‘true’, ‘MEANS’, and ‘THAT’, this sentence has the same status as a logical truth – it is guaranteed to be undeniable by the meanings of these words, along with those of the logical connectives.

## 6.2 the liar

So far, I think, so attractive – we’ve offered a semantics for ‘true’ on the basis of which we can derive the concrete prediction that every instance of the T-schema is undeniable – that it has the very same status as logical truths, taking into account the meanings of ‘true’, ‘MEANS’, and ‘THAT’. In contrast to truth-conditional approaches to the semantics of the truth-predicate, this approach does not need to incorporate any restrictions on the T-schema whatsoever. But does this unrestricted result about the T-schema lead to problems with liar sentences? That is the question to which I now turn.

Our sentential truth predicate allows us to construct the simplest case of a liar-paradoxical sentence:

**liar**       $\sim \text{true}(\text{liar})$

Since ‘liar’ is a name for the sentence which appears just above, liar means that  $\sim\text{true}(\text{liar})$ . That is, we should accept the point that we can capture in our object-language by, ‘MEANS(liar,THAT( $\sim\text{true}(\text{liar})$ ))’. This sentence, after all, simply says that the semantic value of liar is identical to the semantic value of ‘ $\sim\text{true}(\text{liar})$ ’ – which, after all, *is* liar. But now notice that this is the antecedent of an instance of T-Schema, all instances of which are undeniable:

$$\text{T-liar MEANS}(\text{liar,THAT}(\sim\text{true}(\text{liar})))\supset(\text{true}(\text{liar})\equiv\sim\text{true}(\text{liar}))$$

Moreover, since *modus ponens* is a relevance-valid inference, we know that anyone who accepts T-liar and also accepts the stipulated background claim, ‘MEANS(liar,THAT( $\sim\text{true}(\text{liar})$ ))’, is committed to the consequent of T-liar – namely, ‘ $\text{true}(\text{liar})\equiv\sim\text{true}(\text{liar})$ ’. But this is a contradiction. And that, of course, is the seemingly paradoxical nature of the liar. Some obvious background facts about liar sentences appear to commit us to contradictory conclusions.

To avoid these conclusions, many theorists require restrictions on the application of the T-schema. Meanwhile, other theorists have noted that the T-schema exerts ‘pull’ even in those cases which lead to liar-paradoxical conclusions. That sometimes leads them to conclude that natural languages like English, which contain their own truth-predicate, are ‘inconsistent’, in some sense – either that they contain true contradictions, or that their semantic rules are jointly inconsistent, or that semantic competence requires being disposed to accept some contradictions. Biforcated attitude semantics offers a simpler picture of what is going on.

On the picture of biforcated attitude semantics, the reason why all instances of T-Schema exert ‘pull’ is that every instance of T-Schema is a ‘theorem’, in our technical sense of being undeniable without rational inconsistency. But like all theorems in biforcated attitude semantics, even an undeniable sentence may be one that we do not go on to accept. On the contrary, biforcated attitude semantics easily yields the prediction not only that liar and T-liar are both undeniable, but that their negations are both undeniable, as well. So both liar and T-liar are sentences to which the only rationally acceptable attitude is disacceptance full-stop.<sup>14</sup>

This account successfully explains the ‘pull’ of the liar paradox, but without requiring that our language licenses true contradictions, without requiring that the semantic rules of the language are in any way inconsistent, and without requiring that semantic competence requires the

---

<sup>14</sup> See appendix A.3 for the proof.

disposition to accept any contradiction. What semantic competence should naturally be thought of as requiring, on this account, is the appreciation that undeniable premises sometimes lead to contradictions. This does not, however, require accepting those contradictions, because even undeniable premises need not always be accepted. In fact, undeniable premises which entail contradictions are also, ipso facto, *unacceptable* ones. So the only rational course to take is to disaccept them full-stop.

Recall that given the structure of biforcated attitudes, disacceptance full-stop is always a third available option. Moreover, for any sentence in biforcated attitude semantics, it is never rationally inconsistent to disaccept that sentence. That is because every sentence in biforcated attitude semantics expresses a pair of states of being for, one of whose contents is *strictly* stronger than the other. So if ‘P’ expresses  $\langle \text{FOR}(\pi^1)^*, \text{FOR}(\pi^2) \rangle$ , then disaccepting ‘P’ full-stop involves being for each of  $\pi^2$  and  $\neg\pi^1$ . But the assumption that ‘P’ expressed a biforcated attitude is equivalent to the assumption that  $\pi^1$  is strictly stronger than  $\pi^2$ , which is in turn equivalent to the assumption that  $\pi^2$  and  $\neg\pi^1$  are consistent. And since states of being for are inconsistency-transmitting, which means that they disagree with one another just in case their contents are inconsistent, it follows that there is never any rational conflict involved in disaccepting any sentence full-stop. So in particular, there can never be any inconsistency in disaccepting any paradoxical sentence full-stop.

This, I think, amounts to an attractive treatment of liar-paradoxical cases, which generalizes immediately to any paradox with a similar structure. The semantic rules of the language may guarantee that certain sentences are undeniable, and it may be that accepting those sentences commits us to contradictions. But the very fact that this is so guarantees that those very sentences are also *unacceptable*, which means that the only appropriate attitude to have toward them is disacceptance full-stop. Biforcated attitude semantics not only allows for such a third attitude, but allows us to predict and explain when and why it is warranted, on the basis of the assumption that being for is inconsistency-transmitting. And it does this in a semantic framework whose account of ‘true’ is straightforwardly simple, perfectly general, without needing special restrictions or qualifications, perfectly consistent, and which nevertheless explains the ‘pull’ of even paradox-inducing instances of T-Schema: they, too, are guaranteed by the meaning of ‘true’ to be undeniable, and for the very same reasons as the acceptable instances.

## 7 conclusion

In this paper I've tried to present, in as concise a way as I could, the semantic framework of biforcated attitude semantics, and to explain why it constitutes an attractive and general framework for nondescriptivist semantics. Within that framework, I gave an expressivist semantics for the truth predicate, and consequently showed how truth can benefit expressivism, by making good on expressivists' promise to 'earn the right' to truth. And I finished by showing how this kind of expressivist account of truth is also to the benefit of truth, by showing how it enables a perfectly general, very simple, semantics for 'true' which explains the 'pull' of the liar paradox without in any way countenancing the conclusion that natural languages involve any kind of inconsistency. Though they involve some sentences which are both undeniable and unacceptable, there is always an appropriate third response – disacceptance full-stop.

It is essential to this solution, I think, that this account of truth was nondescriptivist – that it endorsed the view that truth is not a property. This is because even if in addition to acceptance and denial, there is a third attitude of disacceptance full-stop, the idea that we could solve the paradox of the liar by simply choosing to neither accept nor deny 'liar is true' looks like an exercise in futility. If truth is a property, and hence corresponds to some way of 'carving up' the world, our refusal to believe either way on how it is carved up doesn't keep it from being carved up one way or the other. Moreover, even the proposal that the world is only indeterminately carved up, or that some properties are only partially determined, leaving space between their extensions and their anti-extensions, only puts off the problem, and leaves us waiting for a 'revenge' version of the paradox.

But if truth is not a property, then in declining to either accept 'liar is true' or deny it, there may be nothing that we are missing out on. Hence, it may not simply be an exercise in futility to elect to disaccept this sentence, along with 'liar is not true', 'liar is true or liar is not true', and 'if liar means that liar is not true, then liar is true iff liar is not true'. In addition to being the only rationally consistent course, there may be nothing to be said for doing otherwise. The expressivist account of truth developed here places meat on the bones of this idea.<sup>15</sup>

---

<sup>15</sup> I owe the main ideas visible in this paper to Barry Lam, who I hope does not object too strenuously to my developing them in this way. Special thanks also to Mike McGlone, to Nikolaj Jang Pederson and Cory Wright, to Matti Eklund and Alexis Burgess, and to Jake Ross, Billy Dunaway, and Scott Soames.

## Appendix A.1

To prove: that ‘TRUE(THAT(P))≡P’ is an  $\text{LU}\{\text{‘TRUE’},\text{‘THAT’}\}$ -theorem. What we want to show, is that for any value of ‘P’, ‘TRUE(THAT(P))≡P’ is undeniable – that is, that the conjunction of the properties in the semantic value of ‘ $\sim(\text{TRUE}(\text{THAT}(\text{P}))\equiv\text{P})$ ’ is a contradiction. First, given the definition of ‘≡’, we can rewrite ‘ $\sim(\text{TRUE}(\text{THAT}(\text{P}))\equiv\text{P})$ ’ as ‘ $\sim(\sim(\text{TRUE}(\text{THAT}(\text{P}))\&\sim\text{P})\&\sim(\text{P}\&\sim\text{TRUE}(\text{THAT}(\text{P}))))$ ’. Assuming, without loss of generality, that the semantic value of ‘P’ is  $\langle\pi^1,\pi^2\rangle$ , our semantics assigns this sentence to semantic value (writing ‘ $\alpha>\beta$ ’ for ‘ $\neg(\alpha\wedge\neg\beta)$ ’):

$$\langle\lambda z(\neg((\text{I}^{\text{major}}(z,\langle\pi^1,\pi^2\rangle)>\pi^2)\wedge(\pi^1>\text{I}^{\text{minor}}(z,\langle\pi^1,\pi^2\rangle))))\rangle^*, \\ \lambda z(\neg((\text{I}^{\text{minor}}(z,\langle\pi^1,\pi^2\rangle)>\pi^1)\wedge(\pi^2>\text{I}^{\text{major}}(z,\langle\pi^1,\pi^2\rangle))))\rangle$$

The conjunction of these two properties is therefore:

$$\neg((\text{I}^{\text{major}}(z,\langle\pi^1,\pi^2\rangle)>\pi^2)\wedge(\pi^1>\text{I}^{\text{minor}}(z,\langle\pi^1,\pi^2\rangle)))\wedge \\ \neg((\text{I}^{\text{minor}}(z,\langle\pi^1,\pi^2\rangle)>\pi^1)\wedge(\pi^2>\text{I}^{\text{major}}(z,\langle\pi^1,\pi^2\rangle)))$$

But our definition of ‘ $\text{I}^{\text{major}}$ ’ and ‘ $\text{I}^{\text{minor}}$ ’ guarantee that this is equivalent to

$$\neg((\pi^1>\pi^2)\wedge(\pi^1>\pi^2))\wedge\neg((\pi^2>\pi^1)\wedge(\pi^2>\pi^1))$$

Which is inconsistent with the assumption that  $\pi^1>\pi^2$  is necessary – which follows from the fact that  $\langle\pi^1,\pi^2\rangle$  is the semantic value of ‘P’.

## Appendix A.2

To prove: that ‘MEANS(s,THAT(P)) $\supset$ (true(s)≡P)’ is an  $\text{LU}\{\text{‘true’},\text{‘MEANS’},\text{‘THAT’}\}$ -theorem. What we want to show, is that for any values of ‘s’ and ‘P’, ‘MEANS(s,THAT(P)) $\supset$ (true(s)≡P)’ is undeniable – that is, that the conjunction of the properties in the semantic value of its negation is contradictory. First, given our definitions for ‘ $\supset$ ’ and for ‘≡’, we re-write its negation as follows: ‘ $\sim(\sim(\text{MEANS}(s,\text{THAT}(\text{P}))\&\sim((\sim(\text{true}(s)\&\sim\text{P})\&\sim(\text{P}\&\sim\text{true}(s))))))$ ’. Assuming, without loss of generality, that the semantic value of ‘P’ is  $\langle\pi^1,\pi^2\rangle$ , our semantics assigns this sentence the semantic value:

$$\begin{aligned}
& \langle \lambda z(\neg(\text{pai}(z, \text{that}(\text{sv}(s) = \langle \pi^1, \pi^2 \rangle))) > ((\text{I}^{\text{major}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(s) = y)))) > \pi^2)) \wedge \\
& (\pi^1 > \text{I}^{\text{minor}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(s) = y)))))) \rangle^*, \\
& \lambda z(\neg(\neg \text{pai}(z, \text{that}(\neg \text{sv}(s) = \langle \pi^1, \pi^2 \rangle))) > ((\text{I}^{\text{minor}}(z, \text{the } \\
& y: (\neg \text{pai}(z, \text{that}(\neg \text{sv}(s) = y)))) > \pi^1)) \wedge (\pi^2 > \text{I}^{\text{major}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(s) = y)))))) \rangle
\end{aligned}$$

And the conjunction of these properties is:

$$\begin{aligned}
& \text{pai}(z, \text{that}(\text{sv}(s) = \langle \pi^1, \pi^2 \rangle)) \wedge \neg((\text{I}^{\text{major}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(s) = y)))) > \pi^2)) \wedge \\
& (\pi^1 > \text{I}^{\text{minor}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(s) = y)))))) \wedge \\
& \neg \text{pai}(z, \text{that}(\neg \text{sv}(s) = \langle \pi^1, \pi^2 \rangle)) \wedge \neg((\text{I}^{\text{minor}}(z, \text{the } y: (\neg \text{pai}(z, \text{that}(\neg \text{sv}(s) = y)))) > \pi^1)) \\
& \wedge (\pi^2 > \text{I}^{\text{major}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(s) = y))))))
\end{aligned}$$

But if  $\text{pai}(z, \text{that}(\text{sv}(s) = \langle \pi^1, \pi^2 \rangle))$ , then the  $y: (\text{pai}(z, \text{that}(\text{sv}(s) = y)))$  is  $\langle \pi^1, \pi^2 \rangle$ , and so on the assumption of the first conjunct, the second conjunct is equivalent to  $\neg((\text{I}^{\text{major}}(z, \langle \pi^1, \pi^2 \rangle) > \pi^2)) \wedge (\pi^1 > \text{I}^{\text{minor}}(z, \langle \pi^1, \pi^2 \rangle))$ . But for the same reasons as in Appendix A.I, given the definitions of ‘ $\text{I}^{\text{major}}$ ’ and ‘ $\text{I}^{\text{minor}}$ ’, this is impossible.

### Appendix A.3

To prove: that disacceptance full-stop of liar and of T-liar is both a consistent, and the only consistent, attitude to have toward those sentences, given that one accepts the background assumption:

$$\mathbf{B} \quad \text{MEANS}(\text{liar}, \text{THAT}(\sim \text{true}(\text{liar})))$$

I’ll first show that given that one accepts B, it is inconsistent to either accept or deny T-liar, and inconsistent to accept or deny liar. Then I’ll show that disaccepting either full-stop is consistent with accepting B, thus yielding the result that disacceptance full-stop is both a consistent, and the only consistent, attitude to have toward those sentences.

To make things simpler, I’ll first simplify the consequent of T-liar:

$$\begin{aligned}
\mathbf{T-liar-cons} \quad & \text{true}(\text{liar}) \equiv \sim \text{true}(\text{liar}) \\
= (\text{defn}) \quad & (\text{true}(\text{liar}) \supset \sim \text{true}(\text{liar})) \& (\sim \text{true}(\text{liar}) \supset \text{true}(\text{liar})) \\
= (\text{defn}) \quad & \sim(\text{true}(\text{liar}) \& \sim \sim \text{true}(\text{liar})) \& \sim(\sim \text{true}(\text{liar}) \& \sim \text{true}(\text{liar}))
\end{aligned}$$

So simplifying, if ‘true(liar)’ has the semantic value  $\langle \lambda z(\pi^1)^*, \lambda z(\pi^2) \rangle$ , (so that the semantic value of liar is  $\langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle$ ) then T-liar-cons has the semantic value:

$$\langle \lambda z(\neg(\pi^2 \wedge \neg\neg\pi^2) \wedge \neg(\neg\pi^1 \wedge \neg\pi^1))^*, \lambda z(\neg(\pi^1 \wedge \neg\neg\pi^1) \wedge \neg(\neg\pi^2 \wedge \neg\pi^2)) \rangle$$

Which is equivalent to the following:

$$\langle \lambda z(\neg\pi^2 \wedge \pi^1)^*, \lambda z(\neg\pi^1 \wedge \pi^2) \rangle$$

Hence I’ll abbreviate it as such in what follows. Modulo this equivalence, then, the semantic value of T-liar is:

$$\langle \lambda z(\neg\text{pai}(z, \text{that}(\neg\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^2 \wedge \pi^1))^*, \lambda z(\text{pai}(z, \text{that}(\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^1 \wedge \pi^2)) \rangle$$

So the semantic value of the negation of T-liar is:

$$\langle \lambda z(\neg(\text{pai}(z, \text{that}(\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^1 \wedge \pi^2))^*, \lambda z(\neg(\neg\text{pai}(z, \text{that}(\neg\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^2 \wedge \pi^1))) \rangle$$

And finally, the semantic value of B is:

$$\langle \lambda z(\text{pai}(z, \text{that}(\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^2 \wedge \pi^1))^*, \lambda z(\neg\text{pai}(z, \text{that}(\neg\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^2 \wedge \pi^1))) \rangle$$

So someone who accepts both B and T-liar is for both properties in the semantic value of B and for both properties in the semantic value of T-liar – but the minor property of the semantic value of B is the antecedent of the major property of the semantic value of T-liar, whose consequent is inconsistent, because ‘true(liar)’ must express a biforcated attitude. Hence, both accepting B and accepting T-liar is inconsistent. We already know that accepting B and denying T-liar is inconsistent, because in Appendix A.2 we proved that any instance of T-schema is inconsistent to deny, and T-liar is an instance of T-schema. So someone who accepts B cannot consistently either accept T-liar or deny it.

Now to disaccept T-liar full stop is to be in the minor attitude of both T-liar and its negation – that is, to be for each of the weaker properties in the semantic values of T-liar and of its negation. We know from above that those properties are

$$\begin{aligned} & \lambda z(\text{pai}(z, \text{that}(\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^1 \wedge \pi^2)) \\ & \text{and} \\ & \lambda z(\neg(\neg\text{pai}(z, \text{that}(\neg\text{sv}(\text{liar}) = \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle)) > (\neg\pi^2 \wedge \pi^1))) \end{aligned}$$

respectively. Someone who both accepts B and disaccepts T-liar full-stop is therefore committed to being for  $\neg\pi^1 \wedge \pi^2$ , because the major property of the semantic value of B is the antecedent of the minor property of the semantic value of T-liar. So she is committed to being for each of  $\neg\pi^1$  and  $\neg\neg\pi^2$ , which, since we have abbreviated by assuming that the semantic value of liar is  $\langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle$ , is what it is to disaccept liar full-stop.

We know, moreover (from section 6.2), that for any given sentence, it is always consistent to disaccept it full-stop. But to show that it is consistent for someone who accepts B to disaccept T-liar and liar full-stop, we have to show that these attitudes are consistent with accepting B – not just that they are self-consistent. It is trivial that disaccepting the negation of T-liar is consistent with accepting B, because the minor property of the semantic value of the negation of T-liar is a trivial property, since we know that  $\pi^1$  entails  $\pi^2$ . Moreover, it is sufficient to show that disaccepting T-liar is consistent with accepting B, to show that disaccepting liar full-stop is consistent with accepting B, since these two states commit to disaccepting liar full-stop.

To evaluate that, we need to remind ourselves what  $\neg\pi^1$  and  $\neg\neg\pi^2$  are.  $\neg\pi^1$  is the property,  $\lambda z(\neg I^{\text{major}}(z, \text{the } y: (\text{pai}(z, \text{that}(\text{sv}(\text{liar}) = y))))$ ), and  $\neg\neg\pi^2$  is the property,  $\lambda z(\neg\neg I^{\text{minor}}(z, \text{the } y: (\neg\text{pai}(z, \text{that}(\neg\text{sv}(\text{liar}) = y))))$ ). So given the properties in the semantic value of B, these commit us to  $\lambda z(\neg I^{\text{major}}(z, \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle))$  and  $\lambda z(\neg\neg I^{\text{minor}}(z, \langle \lambda z(\neg\pi^2)^*, \lambda z(\neg\pi^1) \rangle))$ , respectively. But given the definitions of  $I^{\text{major}}$  and  $I^{\text{minor}}$ , these are equivalent to  $\lambda z(\neg\neg\pi^2)$  and  $\lambda z(\neg\neg\neg\pi^1)$ , respectively – i.e, equivalent to  $\neg\pi^1$  and  $\pi^2$ . No where do we find any inconsistency with accepting B.

## references

Blackburn, Simon [1984]. *Spreading the Word*. Oxford: Oxford University Press.

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

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