Semantic Innocence and Interpreted Logical Forms*

Andrew Botterell†

1. According to Donald Davidson, a semantic theory for a language *L* is *innocent* just in case the semantic values of expressions of *L* do not vary depending upon what environment those expressions are embedded in.[1] A semantic theory for a language *L* is *compositional* just in case the semantic values of complex expressions of *L* are functions of the semantic values of their constituent parts and the way those parts are put together.[2] Call a semantic theory *Davidsonian* just in case it is both innocent and compositional. The idea that a semantic theory should be Davidsonian is an attractive one, since such a theory seems ideally suited to explain the ability of speakers of a language to produce and understand a potential infinity of meaningful expressions. Nonetheless, it is generally assumed that the conjunction of innocence and compositionality places severe constraints on a semantic theory. In particular, it is generally assumed that propositional attitude ascriptions cannot be given an adequate treatment within a Davidsonian semantic theory.

This general assumption has recently been challenged. A number of authors[3] have argued that propositional attitude ascriptions can be given an adequate treatment within a Davidsonian semantic theory if Interpreted Logical Forms (ILFs), or objects similar to ILFs, are taken to be

---

* After years of trying to get this paper published I've more or less given up, although I still believe the paper's basic point is correct. Comments, etc., are still welcome, however.

† Department of Philosophy and Faculty of Law, University of Western Ontario.

[1] See in particular Davidson (1967, 1968, 1979).

[2] Davidson does not use the term 'compositional' to describe the desideratum that the semantic values of complex expressions be functions of the semantic values of their constituent parts. Rather, he requires that a semantic theory be *recursive*, where a semantic theory *T* for a natural language *L* is recursive just in case *T* contains finitely many primitive axioms and combinatorial rules from which can be derived interpretive T-theorems for complex expressions of *L*. However, I will use 'compositional' and its cognates in what follows.

[3] See in particular Higginbotham (1986), Segal (1989), Richard (1990), Larson and Ludlow (1993), Pietroski (1994), (1996), and Segal and Larson (1995). Another important early paper is Burdick (1982), which presents a theory that is similar in important respects to the ILF theories mentioned above.

the objects of propositional attitudes.[4] In Paul Pietroski's opinion, ILF approaches show "how there *could* be a language *L* such that: there is an extensional compositional semantics for *L*, even though some (nonquotational) constructions of *L* are opaque. *So the mere fact that natural languages exhibit opacity in no way shows that such languages do not have a 'Davidsonian' semantics*." (Pietroski 1994, 7 (italics in the original))[5]

Despite the optimism of these ILF theorists, however, it remains unclear exactly how ILF approaches fit in with Davidson's overall semantic program. Indeed, on a natural understanding of what the principles of innocence and compositionality amount to it appears that semantic theories that appeal to ILFs *cannot* be both innocent and compositional. What I wish to do in this paper, then, is look at the claim that ILF approaches preserve what Davidson (1968) has called 'a pre-Fregean semantic innocence' within a compositional semantic theory. I will argue that ILF approaches are innocent only in an attenuated sense. And this, I think, raises a serious problem for an approach to the semantics of propositional attitudes that purports to be, in broad outlines at least, Davidsonian.

The paper proceeds as follows. I begin with a discussion of the constraint of semantic innocence. I then consider the problems attitude ascriptions raise for Davidsonian semantic theories, and describe Davidson's own paratactic solution to those problems. Next, I present a simplified version of Higginbotham and Pietroski's ILF accounts of opacity and argue that they are not semantically innocent in the required sense. Finally, focusing on the ILF theory of Larson and Ludlow, I argue that the problems that infect Higginbotham and Pietroski's ILF accounts

[4] For example, in the case of a belief ascription of the form 'Mary believes that Twain wrote *Huck Finn*' some advocates of the ILF approach take the relevant propositional object to be the ILF denoted by 'that Twain wrote *Huck Finn*', while others take the appropriate propositional object to be something *similar to* the ILF denoted by 'that Twain wrote *Huck Finn*'. Segal (1989), Larson and Ludlow (1993), and Segal and Larson (1995) adopt the former approach; Higginbotham (1986) and Pietroski (1994) the latter. I will ignore these subtleties in what follows.

[5] I should note that Pietroski (1994) does not represent Pietroski's current thinking about opacity; that is to be found in Pietroski (1996). However, since Pietroski (1994) does represent a particularly clear statement of the sort of view that I wish to focus on, I will refer to it in what follows.

infect ILF approaches generally. I conclude with some remarks about the problem of propositional attitude ascriptions and the place of ILFs within it.

2. Let me begin by discussing the constraint of semantic innocence in more detail. Davidson first introduced the constraint of semantic innocence in "On Saying That". There Davidson remarks that

> [s]ince Frege, philosophers have become hardened to the idea that content-sentences in talk about propositional attitudes may strangely refer to such entities as intensions, propositions, sentences, utterances, and inscriptions. What is strange is not the entities, which are all right in their place (if they have one), but the notion that ordinary words for planets, people, tables and hippopotami in indirect discourse may give up these pedestrian references for the exotica. If we could recover our pre-Fregean semantic innocence, I think it would seem to us plainly incredible that the worlds 'The earth moves', uttered after the words 'Galileo said that,' can mean anything different, or refer to anything else, than is their wont when they come in other environments. No doubt their role in *oratio obliqua* is in some sense special; but that is another story. Language is the instrument it is because the same expression, with semantic features (meaning) unchanged, can serve countless purposes." (Davidson 1968, 144-5)

I said above that a semantic theory for a language *L* is semantically innocent just in case the semantic value of expressions of *L* do not vary depending upon what environment those expressions are embedded in. Less formally, a semantic theory is innocent if it does not permit semantic value shifting.[6] By way of illustration, consider Fregean semantic theories.[7] Fregean semantic theories are *not* innocent, since they permit semantic value shifting. For according to Frege, words in 'indirect quotation' "have their indirect reference coinciding with what is customarily their sense." (Frege 1892, 66). So according to Frege the semantic value of 'Twain' in ordinary, extensional contexts is Twain, whereas within the context of an attitude ascription the semantic value of 'Twain' is the *sense* of 'Twain'. Nonetheless, although they are not

[6] I ignore indexicals for simplicity.
[7] See, for example, Frege (1892) and Dummett (1973).

innocent, Fregean theories of attitude ascriptions are compositional. For the semantic value of 'Twain wrote *Huck Finn*', as it occurs in 'Mary believes that Twain wrote *Huck Finn*', which for Frege is a *thought*, will be a function of the semantic values of its constituent parts, which for Frege are *senses*.

According to Davidson, then, Davidsonian semantic theories and Fregean semantic theories differ in that the former are, while the latter are not, semantically innocent. Moreover, this is not an insignificant difference. On the contrary, it is one of the central differences between Davidsonian and Fregean semantic theories, at least as Davidson conceives of things. The following constraint is therefore suggested: any definition of semantic innocence must be capable of distinguishing Fregean semantic theories from Davidsonian semantic theories. Call this the *constraint of nontriviality*. According to the constraint of nontriviality, a definition of semantic innocence that fails to distinguish Davidsonian semantic theories from Fregean semantic theories is not a definition of innocence worth preserving. I will return to discussion of this constraint in Section 5 below.

3. I turn now to the problem attitude ascriptions raise for Davidsonian semantic theories. In order to do so, however, I need to introduce some terminology. For present purposes I will follow the ILF theorists in talking about syntactic forms, where a syntactic form is a *phrase structure marker* of the kind talked about in Chomskian theories of syntax.[8] A phrase structure marker is a labeled tree, i.e., a collection of *points* or *nodes* that stand in various relations of *dominance*.[9] Nodes that dominate other nodes are called *nonterminal* nodes; nodes that dominate

[8] See, for example, Chomsky (1981).
[9] For a more detailed explanation of these and other notions, see Larson and Segal (1995), Chpt. 3.

no other nodes are called *terminal* nodes. I will assume that such phrase structure markers are appropriate structural descriptions of natural language expressions.

A semantic theory is a theory that assigns semantic values to syntactic forms. Letting Val be a function from syntactic forms to semantic values, 'Val($[_{\alpha}...]$)' will refer to the semantic value of the node $[_{\alpha}...]$.[10] The following is an adequate theory for the fragment of English containing only 'Twain', 'Clemens', 'wrote', and '*Huck Finn*':

Primitive Assignments

(A1) Val($[_{NP}$Twain]) = Twain

(A2) Val($[_{NP}$Clemens]) = Twain

(A3) Val($[_{NP}$*Huck Finn*]) = *Huck Finn*

(A4) Val($[_{V}$wrote]) = {<x,y>: x wrote y}

Combinatorial Rules

(C1) Val($[_{S}[_{NP}...][_{VP}[_{V}...]]]$) = *true* iff Val($[_{NP}...]$) ∈ Val($[_{VP}...]$)

(C2) Val($[_{VP}[_{V}$ ...$][_{NP}$ ...$]]]$) = {x: ∃y (y = Val($[_{NP}...]$) & <x,y> ∈ Val($[_{V}...]$))}

Using these rules and standard patterns of inference we can prove the following T-theorems:

(T1) Val($[_{S}[_{NP}$Twain$][_{VP}[_{V}$wrote$][_{NP}$*Huck Finn*$]]]$) = *true* iff Twain ∈ {x: ∃y (y = *Huck Finn* & <x, y> ∈ {<z,w>: z wrote w})}

[10] A contextual parameter should also be included in the domain of the function Val. I omit such a parameter for simplicity.

(T2) Val([$_S$[$_{NP}$Clemens][$_{VP}$[$_V$wrote][$_{NP}$*Huck Finn*]]]) = *true* iff Twain ∈ {x: ∃y (y = *Huck Finn* & <x, y> ∈ {<z,w>: z wrote w})}

As I indicated above, according to Davidson a semantic theory should be both innocent and compositional, and indeed, the simple theory outlined is both. It is compositional because the semantic values of nonterminal nodes are functions of the semantic values of their constituent parts. And it is innocent because nodes are assigned unique semantic values; there is no semantic value shifting. More formally, our simple theory satisfies the following definition of innocence:

Semantic Innocence

A semantic theory *T* is *innocent* iff for all nodes [$_\alpha$...], if according to *T* Val([$_\alpha$...]) = x and Val([$_\alpha$...]) = y, then x = y.[11]

This definition does, I think, capture the desired notion of semantic innocence. It also satisfies the constraint of nontriviality. For Fregean theories do, while Davidsonian theories do not, violate innocence so defined, since according to Fregean theories, sometimes Val([$_{NP}$Twain]) = Twain and sometimes Val([$_{NP}$Twain]) = the *sense* of 'Twain', whereas according to Davidsonian theories, Val([$_{NP}$Twain]) = Twain regardless of where [$_{NP}$Twain] is embedded. In a way, this should come as no surprise. For since Fregean semantic theories permit semantic value shifting, they cannot be formalized using a one-place Val function. It is therefore no surprise to be told that Fregean semantic theories violate Semantic Innocence thus defined.

However, while the simple Davidsonian theory just outlined is adequate for the restricted fragment of English under consideration, it fails for fragments of English that contain verbs of

[11] Since I don't want to make ambiguous languages automatically non-innocent I will assume that the definition of innocence presented above does not apply to such languages, or at least does not apply to ambiguous words occurring in such languages. For example, I don't want to make English automatically non-innocent simply because sometimes Val([$_{NP}$bank]) = financial institution and sometimes Val([$_{NP}$bank]) = side of a river.

propositional attitude. To see why, let us consider a fragment of English which contains 'believes', 'Mary' and 'that' in addition to 'Twain', 'Clemens', 'wrote', and '*Huck Finn*', and let us add the following rules to our simple Davidsonian semantic theory:

Primitive Assignments

(A5) $Val([_{V}believes]) = \{<x,y>: x \text{ believes } y\}$

(A6) $Val([_{NP}Mary]) = Mary$

Combinatorial Rules

(C3) $Val([_{CP}[_{C}that][_{S}...]]) = Val([_{S}...])$

Now consider sentences $[_{S1}...]$ and $[_{S2}...]$:

$[_{S1}[_{NP}Mary][_{VP}[_{V}believes][_{CP}[_{C}that][_{S3}Twain \text{ wrote } \textit{Huck Finn}]]]]$

$[_{S2}[_{NP}Mary][_{VP}[_{V}believes][_{CP}[_{C}that][_{S4}Clemens \text{ wrote } \textit{Huck Finn}]]]]$

According to our semantic theory, the embedded sentences $[_{S3}$Twain wrote *Huck Finn*] and $[_{S4}$Clemens wrote *Huck Finn*] have the same semantic value: each will be true just in case Twain $\in \{x: \exists y (y = \textit{Huck Finn} \ \& \ <x, y> \in \{<z,w>: z \text{ wrote } w\})\}$, i.e., just in case Twain wrote *Huck Finn*. Moreover, if $Val([_{S3}...]) = Val([_{S4}...])$, then according to our semantic theory $Val([_{S1}...]) = Val([_{S2}...])$. But, notoriously, this is not the case. For from the fact that <Mary, $Val([_{CP}[_{C}that][_{S3}...]])> \in \{<x,y>: x \text{ believes } y\}$ it does not follow that <Mary, $Val([_{CP}[_{C}that][_{S4}...]])> \in \{<x,y>: x \text{ believes } y\}$. In short, our simple semantic theory makes the wrong predictions about the truth conditions of attitude ascriptions.

4. In response to the problem raised by attitude ascriptions, Davidson (1968) proposed his *paratactic* analysis of attitude ascriptions.[12] According to Davidson, a sentence like [$_{S1}$...] is really composed of two sentences, viz., 'Mary believes that.' and 'Twain wrote *Huck Finn*.' Davidson suggests that 'that' in [$_{S1}$...] functions as a demonstrative that picks out an utterance or inscription of the sentential form immediately following it. Appealing to a primitive notion of *same-saying*, Davidson suggests that [$_{S1}$...] will be true just in case there exists an utterance (or inscription) *x* such that Mary believes *x* and *x* and my next utterance (or inscription) make us same-sayers. Twain wrote *Huck Finn*.

Davidson's paratactic theory has the virtue of being both compositional and innocent. It is compositional because the semantic value of [$_{S1}$...] is a function of the semantic value of [$_{S3}$...]; and the semantic value of [$_{S3}$...] is a functions of the semantic values of its constituent parts. It is innocent because the semantic values of terminal and nonterminal nodes remain fixed. Moreover, Davidson's paratactic analysis does seem to make the correct predictions about the semantic values of [$_{S1}$...] and [$_{S2}$...]. Since according to Davidson the objects of belief are utterances or inscriptions, Mary can stand in the same-saying relation to the utterance or inscription [$_{S3}$...] without standing in the same-saying relation to the utterance or inscription [$_{S4}$...]. Consequently, Val([$_{S1}$...]) can be *true* even though Val([$_{S2}$...]) is *false*. In short, Davidson accounts for the fact that Val([$_{S1}$...]) is not identical to Val([$_{S2}$...]) by rejecting the standard syntactic analysis of attitude ascriptions.

If Davidson's paratactic theory were successful, appeal to ILFs would be unnecessary. It is generally agreed, however, that Davidson's paratactic analysis of attitude ascriptions is syntactically misguided: contrary to what Davidson suggests, 'that' functions as a complementizer in sentences like [$_{S1}$...], not as a demonstrative.[13] It remains to be seen, however,

[12] Davidson only explicitly discusses indirect discourse in Davidson (1968). He does remark, however, that his paratactic analysis of indirect discourse "opens a lead to an analysis of psychological sentences generally (sentences about propositional attitudes, so-called)[.]" (Davidson 1968, 93) So I don't think I am being unfair in attributing the above proposal to Davidson.

[13] See, for example, Segal and Speas (1986) and Segal (1989).

whether a semantic theory which appeals to ILFs and which accepts the standard syntactic analysis of attitude ascriptions can preserve both compositionality and innocence. With this question in mind, I turn now to Higginbotham and Pietroski's ILF approaches to attitude ascriptions.

5. According to ILF approaches, attitude ascriptions express a relation between an agent and an ILF, where an ILF is a phrase structure marker that encodes semantic as well as syntactic information.[14] To simplify matters, let us say that for all nodes $[_{\alpha}...]$, the ILF of $[_{\alpha}...]$ is $\S[_{\alpha}...]\S$. Thus, for example, $\S[_{S3}...]\S$, the (very much simplified) ILF of $[_{S3}...]$, might be represented as follows:

$[_{S}[_{NP}$<'Twain', Twain>$][_{VP}[_{V}$<'wrote', {<x,y>: x wrote y}>$][_{NP}$<'*Huck Finn*', *Huck Finn*>$]]]$

$\S[_{S3}...]\S$ is obtained from $[_{S3}...]$ by replacing the lexical items in $[_{S3}...]$ with ordered pairs consisting of lexical items and the ordinary semantic values of those lexical items.

Pietroski (1994), following Higginbotham (1986), presents a semantic theory that eschews a one-place Val function in favor of a *two*-place Val function. Again, letting Val be a function from syntactic forms to semantic values, 'Val($[_{\alpha}...]$, $[_{\beta}...]$)' will refer to the semantic value of the node $[_{\alpha}...]$ evaluated relative to the node $[_{\beta}...]$. Pietroski then considers the principle that "the value of a constituent does not depend upon what it is embedded in." (Higginbotham 1986, 33) Higginbotham calls this the *indifference principle*:

[14] Again, note that advocates of similarity, such as Higginbotham and Pietroski, think that propositional attitude ascriptions express a relation between agents and something that is *similar to*, although possibly distinct from, an ILF.

Indifference

A semantic theory *T* is *indifferent* iff for all nodes [$_{\alpha}$...], [$_{\beta}$...], if according to *T* Val([$_{\alpha}$...], [$_{\beta}$...]) = x and Val([$_{\alpha}$...], [$_{\alpha}$...]) = y, then x = y.

According to the indifference principle, the semantic value of a node does not vary depending on what context it is embedded in, or depending on what node it is evaluated relative to. However, the problem raised by attitude ascriptions would seem to indicate that the indifference principle must be rejected.

Since we are now working with a two-place Val function, both our primitive assignments and our list of combinatorial rules must be revised. Pietroski is thus led to replace (C3) with (C4):

(C4) Val([$_{S}$...], [$_{CP}$[$_{C}$that][$_{S}$...]]) = §[$_{S}$...]§

Note, however, that if (C4) is adopted Pietroski is now capable of distinguishing Val([$_{S1}$...], [$_{S1}$...]) from Val([$_{S2}$...], [$_{S2}$...]).[15] Because §[$_{S3}$...]§ and §[$_{S4}$...]§ contain different constituent parts, §[$_{S3}$...]§ and §[$_{S4}$...]§ are different objects. Consequently, it does not follow from the fact that <Mary, §[$_{S3}$...]§> ∈ {<x,y>: x believes y} that <Mary, §[$_{S4}$...]§> ∈ {<x,y>: x believes y}. In short, if (C4) is adopted, Val([$_{S1}$...], [$_{S1}$...]) can be true even though Val([$_{S2}$...], [$_{S2}$...]) is false. And that is precisely the desired result.

Importantly, however, while the adoption of (C4) would seem to resolve the problems raised by attitude ascriptions, it contradicts Pietroski's claim that his and Higginbotham's account of opacity preserve a pre-Fregean semantic innocence. According to Higginbotham and Pietroski, Val([$_{S3}$...], [$_{S3}$...]) is a truth-value. According to Higginbotham and Pietroski, however, Val([$_{S3}$...], [$_{CP}$[$_{C}$that][$_{S3}$...]]) = §[$_{S3}$...]§. But this is just to say that Higginbotham and Pietroski's

[15] Again, recall that Pietroski is working with a two-place Val function.

accounts of opacity permit semantic value shifting. Thus it would appear that Higginbotham and Pietroski's ILF approaches violate our pre-Fregean semantic innocence, at least as Davidson conceives of it.

As against this, however, recall that innocence is formally defined using a one-place Val function, whereas Higginbotham and Pietroski's semantic theories appeal to a two-place Val function. Consequently, since Higginbotham and Pietroski's semantic theories cannot satisfy the antecedent of the conditional in the formal definition of innocence, their semantic theories *trivially* satisfy the constraint of innocence so defined. However, this observation is of little comfort, since the notion of semantic innocence thus preserved does not satisfy the constraint of nontriviality. To see why this is so, consider a Fregean semantic theory that also utilizes a two-place Val function. This Fregean theory will also be innocent since, like Higginbotham and Pietroski's semantic theories, it will fail to satisfy the antecedent of the conditional in the formal definition of innocence. But again, since I take it to be a constraint on a definition of semantic innocence that it be able to distinguish Fregean theories from Davidsonian theories, Higginbotham and Pietroski cannot claim that their theories have the virtue of being innocent in any interesting pre-Fregean sense.

6. To this point I have been primarily concerned with Higginbotham and Pietroski's ILF theories. I have argued that neither theory is semantically innocent in the required sense. I also said, however, that I would argue that the problems that infect Higginbotham and Pietroski's ILF theories infect ILF theories quite generally. So let me turn to a discussion of another ILF approach to propositional attitude ascriptions, that of Richard Larson and Peter Ludlow. Discussion of this ILF theory will pave the way to a more general criticism of ILF theories.

Like Higginbotham and Pietroski, Larson and Ludlow (1993)—hereafter *L&L*—attempt to marry a broadly Davidsonian semantic theory with ILFs. The core of L&L's theory is a recursive theory of material truth along the lines proposed by Tarski. Thus in L&L's semantics, each expression is assigned an extensional semantic value, with the semantic values of complex

expressions being specified in terms of the semantic values of their constituent parts. So, for example, L&L's semantics would endorse the following (somewhat simplified) primitive assignments and combinatorial rules:

Primitive Assignments

(A7) Val(x, [$_{NP}$ Mary]) iff x = Mary

(A8) Val(x, [$_{NP}$ Twain]) iff x = Twain

(A9) Val(x, [$_{NP}$ *Huck Finn*]) iff x = Huck Finn

Combinatorial Rules

(C5) Val(t, [$_{S}$ NP VP]) iff for some x, Val(x, NP) and Val(x, VP)

(C6) Val(x, [$_{VP}$ V]) iff Val(x, V)

(A7) is to be read as saying that x is a value of the [$_{NP}$ Mary] iff x = Mary. And (C5) is to be read as saying that a sentence of the form [$_{S}$ NP VP] will be assigned the value true just in case there is an object x such that x is a value of the NP and x is a value of the VP. So, for example, Val(t, [$_{S}$Mary runs]) iff for some x, Val(x, [$_{NP}$Mary]) and Val(x, [$_{VP}$runs]), i.e., iff for some x, x = Mary and x runs, i.e., iff Mary runs.

L&L introduce ILFs into their theory by means of the following axiom for clause-embedding verbs (where again, for all nodes [$_{\alpha}$...], the ILF of [$_{\alpha}$...] is §[$_{\alpha}$...]§):

(C7) Val(x, [$_{VP}$ V S]) iff for some y, Val(<x, y>, V) and y = §S§.[16]

[16] Following L&L, I omit the complementizer [$_{C}$ that].

Thus, ILFs appear whenever there is a VP which contains a clause-embedding verb V together with a complement S. The semantic axiom for a clause-embedding verb such as 'believes' is as follows:

(A 10) Val(<x,y>, [$_V$ believes]) iff x believes y.

How do L&L construct their ILFs? In much the same way in which Pietroski and Higginbotham construct theirs. Again, simplifying somewhat, the following gives the general inductive definition of the ILF of an expression α:

(C8) Let α be a phrase structure marker with root S, and let β be a sub-phrase structure marker of α. Then if:

(i) β is a terminal node, then §β§ = <β, x>.

(ii) β is a non-terminal node composed of [γ1, γ2, ..., γn], then §β§ = [§γ1§ §γ2§...§γn§].

So if an expression β is a terminal node, then the ILF of β will be an ordered pair consisting of β together with a semantic value assigned to it by Val. And if β is a non-terminal node which has as its immediate constituents the expressions γ1, γ2, ..., γn, then the ILF of β will be an ordered n-tuple composed of the ILFs of each of γ1, γ2, ..., γn, respectively.

Now, consider again our simple sentences (S3) and (S4):

(S3) [$_{S3}$Twain wrote *Huck Finn*]

(S4) [$_{S4}$Clemens wrote *Huck Finn*]

According to L&L's semantics, Val(t, [$_{S3}$Twain wrote *Huck Finn*]) iff for some x, x = Twain and x wrote *Huck Finn*; or more simply, iff Twain wrote *Huck Finn*. Similarly, Val(t, [$_{S3}$Clemens wrote *Huck Finn*]) iff Twain wrote *Huck Finn*. In short, according to L&L's semantics both (S3) and (S4) get assigned truth-values.

However, what happens when (S3) and (S4) appear as complements of clause-embedding verbs, as follows:

(S1) [$_{S1}$[$_{NP}$Mary][$_{VP}$[$_{V}$believes][$_{S3}$Twain wrote *Huck Finn*]]]

(S2) [$_{S2}$[$_{NP}$Mary][$_{VP}$[$_{V}$believes][$_{S4}$Clemens wrote *Huck Finn*]]]

The derived biconditional for (S1) is as follows:

(i) Val(t, [$_{S1}$Mary believes Twain wrote *Huck Finn*]) iff Mary believes §[$_{S3}$Twain wrote *Huck Finn*]§.

In short, Val(t, [$_{S1}$Mary believes Twain wrote *Huck Finn*]) iff Mary believes the ILF of [$_{S3}$Twain wrote *Huck Finn*]. And what is the ILF of [$_{S3}$Twain wrote *Huck Finn*]? It is: [§[$_{NP}$Twain]§ §[$_{V}$wrote]§ §[$_{NP}$Huck Finn]§], i.e., the ordered triple that is the result of pairing each lexical item in [$_{S3}$Twain wrote *Huck Finn*] with its ordinary semantic value.

Moreover, we can easily see that according to these same principles, Val(t, [$_{S1}$Mary believes Clemens wrote *Huck Finn*]) iff Mary believes the ILF of [$_{S4}$ Clemens wrote *Huck Finn*]. And again, since §[$_{S3}$Twain wrote *Huck Finn*]§ is distinct from §[$_{S4}$Clemens wrote *Huck Finn*]§, L&L's theory makes the right predictions about our sentences (S1) and (S2), since it allows that (S1) can be true even though (S2) is false.

So far, so good. However, let us again ask: does this semantic theory preserve semantic innocence? On reflection, it seems clear that it does not. For according to L&L's theory, the semantic value assigned to [$_{S3}$Twain wrote *Huck Finn*] when it is unembedded is a truth-value.

On the other hand, the semantic value assigned to [$_{S3}$Twain wrote *Huck Finn*] when it appears as the complement of a clause-embedding verb is an ILF. So it would appear that semantic value shifting has again occurred, and innocence has been lost. In a way this should come as no surprise. For since L&L employ a binary valuation relation, there is no longer such a thing as *the* semantic value of an expression; there is only the notion of *a* semantic value for an expression evaluated relative to larger expressions in which it is embedded. And this is precisely what the constraint of semantic innocence is designed to rule out.

7. By way of response, it might be objected that the preceding argument trades on a misrepresentation of the way in which L&L's semantic theory is actually constructed, and that once this misrepresentation is cleared up the preceding argument loses its force. For consider L&L's clause for the introduction of ILFs:

(C7) Val(x, [$_{VP}$ V S]) iff for some y, Val(<x, y>, V) and y = §S§.

According to this clause a semantic value is not assigned to S nodes when they occur within the scope of clause embedding verbs; rather, the entire verb phrase receives a semantic value, viz., an ordered pair, the first element of which is an individual and the second element of which is an ILF. So, for example, Smith will be a semantic value of the VP 'believes pigs fly' iff <Smith, §pigs fly§> is a semantic value of 'believes'. And if the S node is not assigned a semantic value when it occurs within the scope of a clause embedding verb, then it is arguable that its semantic value cannot be said to shift from a truth-value to an ILF.

Similarly, recall that Pietroski offered us

(C4) Val([$_{S}$...], [$_{CP}$[$_{C}$that][$_{S}$...]]) = §[$_{S}$...]§

And as we saw, (C4) leads to trouble. But this is easily fixed. For suppose we replace (C4) with (C4*):

(C4*) Val([$_{CP}$[$_{C}$that][$_{S}$...]], [$_{CP}$[$_{C}$that][$_{S}$...]]) = §[$_{S}$...]§

That is, suppose we assign a semantic value not to an embedded S node itself, but rather to the entire complementizer phrase of which the S node is a part. Then by making the ILF the semantic value of the complementizer phrase [$_{CP}$[$_{C}$that][$_{S}$...]], a semantic value is no longer assigned to S, and its semantic value cannot be said to shift.

Attractive as this proposal may be, however, there are two serious problems with it. First, even if it is right to say that the S node does not receive a semantic value when it occurs within the scope of a clause-embedding verb, this does not automatically entail that its semantic value does not shift. Rather, it would be more appropriate to say that its semantic value 'shifts' from a truth-value to no semantic value at all. Second, and more importantly, even if this attempt to preserve semantic innocence were successful, it conflicts with the other part of Davidson's semantic program I mentioned above, namely the constraint of compositionality. Earlier I said that a semantic theory for a language *L* is *compositional* just in case the semantic values of complex expressions of *L* are functions of the semantic values of their constituent parts and the way those parts are put together. But this leads to problems. For consider: if L&L's theory is to be compositional, then in the case of something of the form '[$_{VP}$ V S]' the S node must get assigned *some* semantic value, and the semantic value it gets assigned must combine with the semantic value of the V node to yield a unique semantic value for the VP node. However, suppose we allow that the S node fails to receive a semantic value, and so fails to shift its semantic value. Then the semantic value of the *VP* node in question will not be a function of the semantic values of its constituent parts and the way those parts are put together. In short, the problem with this suggestion is that it either fails to preserve semantic innocence, or it conflicts

with the constraint of compositionality. In neither case is it a move that philosophers concerned with defending a broadly Davidsonian semantic theory ought to make.

8. The problem that consistently recurs is this: there appears to be no way to preserve semantic innocence while endorsing the idea that the semantic value of a complementizer phrase is an ILF. The best that ILF theorists can do, it seems to me, is to adopt something along the following lines.[17] Let us say that the semantic value of a sentence is always and everywhere a truth-value. Let us also say that the semantic value of a complementizer [$_C$ that] as it occurs in a sentence of the form 'X believes that P' is the *uninterpreted* logical form of 'P', i.e., is the uninterpreted logical form of the sentence which the clause embedding verb preceding [$_C$ that] embeds. Then the following proposal is possible: the semantic value of 'that P' is an ordered pair, the first element of which is the semantic value of 'that', and the second element of which is the semantic value of 'P'. But the semantic value of 'that' is the uninterpreted logical form of 'P', and the semantic value of 'P' is a truth-value. So the semantic value of 'that P' is the *interpreted* logical form of 'P'. So a sentence of the form 'X believes that P' will be true just in case X believes the ILF of 'P'. And again, that is precisely the result that ILF theorists are after.

We can make this idea a bit more precise by adopting the following axioms:

(C?) Val([$_C$ that], [$_{CP}$[$_C$that][$_S$...]]) = the (uninterpreted) LF of [$_S$...]

(S?) Val([$_S$...], [$_{CP}$[$_C$that][$_S$...]]) = Val([$_S$...], [$_S$...])

(CS?) Val([$_{CP}$[$_C$that][$_S$...]], [$_{CP}$[$_C$that][$_S$...]]) = §[$_S$...]§

[17] See Rumfitt (1993) and Pietroski (1996) for proposals along similar lines.

Note that this proposal would appear to be both compositional and innocent. It is compositional since the semantic value of a complementizer phrase [$_{CP}$[$_{C}$that][$_{S}$...]] is a function of the semantic values of its constituent parts. And it is innocent since the semantic value of [$_{S}$...] remains a truth-value regardless of where it is embedded. So why not simply adopt this style of proposal and have done with the entire discussion of innocence, compositionality, and ILFs?

Two problems indicate why we should not do this. The main problem with this proposal is that it is unclear what the mechanism is for generating the semantic value of the complementizer [$_{C}$that]. On the current proposal, [$_{C}$that] functions as a sort of indexical or demonstrative: it points to, and hence acquires its reference from, the sentence that follows it. But how is this pointing supposed to work? It is clear that [$_{C}$that] cannot be a true demonstrative, since it functions syntactically as a complementizer. As against this it might be suggested that some sort of transformation occurs at the level of LF, whereby the complementizer becomes a demonstrative.[18] But this is entirely ad hoc. Lastly, perhaps [$_{C}$that] can be viewed as a functional expression that takes an embedded clause as argument and yields and ILF as value. This is a plausible suggestion. But it leads to the second problem. For it again seems that the Fregean can mimic this ILF view and thereby preserve innocence.

To see how, suppose the Fregean suggests the following: let the semantic value of a sentence be a truth-value. But let the semantic value of a complementizer [$_{C}$that] as it occurs in a sentence of the form 'X believes that P' be the *sense* of 'P'. Then we get the following proposal: the semantic value of 'that P' is an ordered pair, the first element of which is the semantic value of 'that', and the second element of which is the semantic value of 'P'. But the semantic value of 'that' is the sense of 'P', and the semantic value of 'P' is a truth-value. The semantic value of 'that P' is thus what we might call the *Fregean* interpreted logical form of 'P', and a sentence of the form 'X believes that P' will be true just in case X believes the Fregean ILF of 'P'.

[18] This is Rumfitt's (1993) suggestion.

Moreover, this Fregean proposal seems to be compositional, and it also appears to be innocent. However, since it violates the constraint of non-triviality, the fact that the proposed ILF theory also satisfies this constraint of semantic innocence is of limited interest.

9. My conclusion at this point is that none of the ILF theories considered thus far is semantically innocent in the required sense. A final complication needs to be addressed before we can conclude, however. For somebody might reasonably respond that even if the semantic value of [$_{S3}$Twain wrote *Huck Finn*] shifts from a truth-value to an ILF when it occurs within the scope of a clause embedding verb, something remains constant in the two cases. In particular, somebody might argue that regardless of where it is embedded [$_{S3}$Twain wrote *Huck Finn*], say, still *refers to* a truth-value, and hence, that [$_{S3}$Twain wrote *Huck Finn*] never 'gives up its pedestrian reference for the exotica'. Similarly, it might be argued that even when [$_{S3}$Twain wrote *Huck Finn*] occurs within the scope of a clause embedding verb, the [$_{NP}$Twain], say, still *refers to* Twain, and so preserves its pedestrian reference. Indeed, it is hard to see how any of the ILF theories considered could fail to entail this, since the recursive definition for building an ILF explicitly incorporates the ordinary semantic value of [$_{\alpha}$...] into the ILF for [$_{\alpha}$...]. Moreover, if the preservation of reference is sufficient to preserve innocence, then ILF theories preserve semantic innocence after all.

This objection gives rise to two important questions. First, how should 'semantic innocence' be defined? And second, how should the expression 'semantic value' be understood? These two questions are not unrelated. In first introducing the constraint of semantic innocence Davidson remarked that "language is the instrument it is because the same expression, with semantic features (meaning) unchanged, can serve countless purposes." (Davidson 1968, 145) Notoriously, however, there are many different ways to make sense of talk about semantic features, or meaning. Talk of intension, extension, sense, reference, use, connotation, and denotation all come to mind. In consequence, it seems to me that we would do well to distinguish at least two ways in which a semantic theory might be said to be semantically innocent. Let us

say that a semantic theory *T* for a natural language *L* is *weakly innocent* if according to *T* the referent of an expression α of *L* does not vary depending on what environment α is embedded in. And let us say that a semantic theory *T* for a natural language *L* is *strongly innocent* if according to *T* the elements which enter into the determination of the semantic value of an expression α of *L* in extensional contexts are identical with the elements which enter into the determination of the semantic value of α in non-extensional contexts. Strong innocence therefore entails weak innocence, but not vice versa.

This distinction is important, since it permits us to argue for a perfectly general conclusion regarding ILF theories. And this conclusion is that *any* ILF theory, no matter how it is articulated, will fail to be strongly innocent. This is because at the core of any ILF theory is the idea that in the scope of a clause-embedding verb like 'believes', an expression takes as its semantic value an ordered pair consisting of the ordinary referent of the expression together with the expression itself. For example, recall L&L's axiom for the construction of ILFs: if β is a terminal node, then if Val(x, β), then §β§ = <β, x>. In other words, to obtain the ILF for an expression α, we simply pair α with its ordinary semantic value, i.e., with its referent. But this means that ILF theorists are committed to thinking that whenever a sentence occurs within the scope of a clause embedding very, its semantic value is enriched with additional semantic features. And it is precisely this enrichment of semantic features that precludes ILF theories from satisfying the constraint of strong semantic innocence.

To be sure, on the ILF approach ordinary extensional reference is always preserved: a sentence always refers to a truth-value, and a (referring) proper name always refers to an individual. But this is of no help if what is wanted is the preservation of an interesting, i.e., strong, notion of semantic innocence. For suppose we grant that weak semantic innocence is preserved if we let Val([$_{NP}$Twain]) = <'Twain', Twain>, and if we stipulate that in extensional contexts the semantics 'sees through' or ignores 'Twain'. Then it must also be granted that the Fregean can preserve weak semantic innocence by insisting that Val([$_{NP}$Twain]) = <the sense of 'Twain', Twain> and by stipulating that in extensional contexts the semantics ignores the sense

of 'Twain'.[19] So if this sort of move preserves weak semantic innocence in the case of ILF theories, it is open to a Fregean to claim that her theory is weakly semantically innocent as well.

For perfectly general reasons, then, it is unclear to me whether ILF theories are capable of preserving an interesting version of semantic innocence.[20]

10. The motivation for introducing ILFs into an extensional semantic theory is easy enough to understand. Binary theories of propositional attitude ascriptions that posit purely syntactic objects as the objects of belief encounter problems when two objects of belief have the same form but different content. Binary theories of propositional attitude ascriptions that posit purely semantic objects as the objects of belief encounter problems when two objects of belief have the same content but different form. Objects that have both syntactic and semantic properties would thus appear to be ideally suited to be the objects of propositional attitudes. Nevertheless, I have argued that in the case of a Davidsonian semantic theory the appearance of progress is illusory. ILF approaches succeed only because they abandon semantic innocence, and with it, one of the constraints of Davidson's general semantic picture. So if semantic innocence is a constraint worth preserving, this suggests that ILF theories should be rejected.

But *is* semantic innocence a constraint worth preserving? I have been assuming that, other things being equal, a semantic theory that is innocent is to be preferred to a semantic theory

[19] Burge (1986) offers the Fregean a proposal along these lines.

[20] This does not preclude the introduction of other definitions of 'semantic innocence.' For example, L&L suggest that "the analysis of propositional attitudes cannot introduce anything new into the ontology [of the semantic theory]; it must be semantically 'innocent' in the sense of Davidson [1968]." (L&L 332) Similarly, Larson and Segal remark that "[a]s simple amalgams of words, phrases, and the objects they refer to, ILFs invoke no entities not already required for giving the semantics of the simplest parts of the grammar. No new, special entities like propositions, relations or properties are introduced. The ILF theory is thus semantically innocent[.]" (Larson and Segal, 441-2) This is, of course, true. But it is not an interesting notion of semantic innocence. For consider a Fregean theory which appeals to *senses* in order to give the semantics for the simplest parts of the grammar, either because the Fregean theory takes the semantic values of terminal nodes to be senses, or because the Fregean theory takes the semantic values of terminal nodes to be determined by the senses of those terminal nodes. According to Larson and Segal's definition of innocence, if such a theory also appealed to senses in order to give a semantics for propositional attitudes it would be semantically innocent. But again, a definition of semantic innocence that does not distinguish Fregean theories from Davidsonian theories is not a definition of semantic innocence that anybody should be interested in.

that is not. But given our distinction between weak and strong innocence this assumption needs to be reconsidered. For while weak innocence is easy enough to preserve, it is uninteresting; and while strong innocence is interesting, it is very hard, if not impossible, to come by. In short, if what is at issue in discussions of semantic innocence is weak innocence, then it should be granted that ILF theories do preserve semantic innocence, and so are semantically innocent. It should also be noted, however, that preserving weak semantic innocence is not a significant accomplishment, since any semantic theory—including Fregean semantic theories—can preserve semantic innocence if it is willing to complicate its primitive axioms and combinatorial rules sufficiently. On the other hand, if what is at issue in discussions of semantic innocence is strong innocence, then I think it must be concluded that ILF theories, as well as Fregean theories, fail to be semantically innocent.

Which constraint of semantic innocence, weak or strong, ought to be preferred? To answer this question we first need to ask whether the plausibility of a semantic theory depends on its according with our pre-theoretical views about reference and meaning. Do we, for instance, think that 'Twain' must everywhere and always refer to Twain, or have Twain as its semantic value? I am inclined to think that our pre-theroretical views about reference and meaning are largely irrelevant to issues about semantic innocence. Semantic innocence is defined in terms of the notion of a semantic value, and that notion has very little to do with ordinary ideas about the meanings of natural language expressions. In this I agree with Tyler Burge. As Burge remarks in a related discussion, "The notion of…[semantic] value is a theoretical extension of the notion of reference [and is] more theoretical than that of reference. There are solid theoretical reasons to think that the [semantic] values of terms in non-extensional contexts differ from those in extensional contexts. I think that intuitions about the reference of terms are of severely limited import to this issue." (Burge 1986, 207) In short, like Burge I think it is a mistake to argue from pre-theoretical, or pre-Fregean, intuitions about reference to conclusions about whether the semantic values of expressions can shift. Does it follow from what I have said that ILF approaches to propositional attitude ascriptions cannot be correct? No. But this is as it

should be. For in the end the important question is not whether our pre-Fregean semantic innocence is worth preserving. Rather, the question is whether ILFs are appropriate objects of propositional attitudes like belief. And it is not clear to me that the constraint of semantic innocence has much bearing on this question.

References

Burdick, H. 1982. "A Logical Form for the Propositional Attitudes." *Synthese* 52, 185-230.

Burge, T. 1986. "On Davidson's 'Saying That'." In E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. (Oxford: Blackwell, 1986), 190-208.

Chomsky, N. 1975. *Reflections on Language*. (New York: Pantheon).

Chomsky, N. 1981. *Lectures on Government and Binding*. (Dordrecht: Foris).

Chomsky, N. 1986. *Knowledge of Language: Its Nature, Origin, and Use*. (New York: Praeger).

Davidson, D. 1967. "Truth and Meaning." In Davidson, D. *Inquiries into Truth and Interpretation*. (Oxford: Oxford University Press, 1984).

Davidson, D. 1968. "On Saying That." In Davidson, D. (1984).

Davidson, D. 1979. "Quotation." In Davidson, D. (1984).

Dummett, M. 1973. *Frege: Philosophy of Language*. (Cambridge, MA: Harvard University Press).

Frege, G. 1892. "On Sense and Reference." In P. Geach and M. Black (eds.) *Translations from the Philosophical Writings of Gottlob Frege*. (Oxford: Basil Blackwell, 1970).

Higginbotham, J. 1986. "Linguistic Theory and Davidson's Program in Semantics." In E. LePore (ed.) *Truth and Interpretation: Perspectives on the Philosophy of Donald Davidson*. (Oxford: Blackwell, 1986), 29-48.

Larson, P. and Ludlow, P. 1993. "Interpreted Logical Forms." *Synthese* 95, 305-55.

Larson, P. and Segal, G. 1995. *Knowledge of Meaning*. (Cambridge, MA: MIT Press).

Oppy, G. 1992. “Why Semantic Innocence?” *Australasian Journal of Philosophy* 70, No. 4, 445-454.

Pietroski, P. 1994. “Similarity and Innocent Semantics.” McGill University Cognitive Science Center, Manuscript #294.

Pietroski, P. 1996. “Fregean Innocence.” *Mind and Language* 11, No. 4, 338-370.

Rumfitt, I. 1993. “Content and Context: The Paratactic Theory Revisited and Revised.” *Mind* 102, 429-54.

Segal, G. 1989. “A Preference for Sense and Reference.” *Journal of Philosophy* 86, 73- 89.

Segal, G. and P. Speas. 1986. “On Saying Qet.” *Mind & Language* 1, No. 2, 124-132.