Variations on a Theme by Yablo

Hartry Field

Naive truth theory is, roughly, the theory of truth that in classical logic leads to well-known paradoxes (such as the Liar paradox and the Curry paradox). One response to these paradoxes is to weaken classical logic by restricting the law of excluded middle and introducing a conditional not defined from the other connectives in the usual way. In "New Grounds for Naive Truth Theory" ([12]), Steve Yablo develops a new version of this response, and cites three respects in which he deems it superior to a version that I've advocated in several papers. I think he's right that my version was non-optimal in some of these respects (one and a half of them, to be precise); however, Yablo's own account seems to me to have some undesirable features as well. In this paper I will explore some variations on his account, and end up tentatively advocating a synthesis of his account and mine (one that is somewhat closer to mine than to his).

1 Background.

First some philosophical motivation for the project that Yablo and I share. A standard account of why we need a notion of truth (Quine [10], Leeds [8]) is to allow us to make and use generalizations not expressible (or not easily expressible) without the notion. For instance, even if I don't remember all the details of what you said yesterday but only the general gist of it, I may say "If everything you said yesterday is true then probably I should change my

plans about such and such"; the antecedent is supposed to be equivalent to the conjunction of everything you said, which I'm not in a position to fill in. It would seem that in order for the notion of truth to fill this role, the attribution of truth to a sentence has to be fully equivalent to the sentence itself: call this the Equivalence Principle. More precisely, $True(\langle A \rangle)$ must be intersubstitutable with A in all extensional contexts, including in the scope of conditionals.[1] (It wouldn't be enough for $True(\langle A \rangle)$ and A to be co-assertable, or for them to be co-assertable and in addition $\neg True(\langle A \rangle)$ and $\neg A$ to be co-assertable; in the example given, intersubstitutability within the antecedent of a conditional is required.)

So the Equivalence Principle is not something we should give up lightly. Unfortunately, it conflicts with classical logic: in languages with very minimal resources for developing syntax we can formulate "liar sentences" which assert their own untruth. Such a sentence L is equivalent to $\neg True(\langle L \rangle)$; but the equivalence principle requires it to also be equivalent to $True(\langle L \rangle)$; so $True(\langle L \rangle)$ must be equivalent to $\neg True(\langle L \rangle)$, which is impossible in classical logic. The problem is not reasonably blamed on the syntactic resources that allow for self-reference: among other reasons for this, there's the fact that the motivation for the Equivalence Principle extends to a similar principle involving the notion of *true of*, and applying this extended principle to the predicate 'is not true of itself' we get a contradiction in classical logic even without the use of any special syntactic assumptions. The only two serious choices are keeping classical logic while restricting the Equivalence Principle, and keeping the Equivalence Principle while restricting classical logic.

The first of these choices has been well explored, on the technical level: see for instance Friedman and Sheard [5] for an account of the main sub-options

[1] Possibly in some non-extensional contexts too, but there will be no need to decide this here. We certainly don't need them to be intersubstitutable in quotational contexts or in propositional attitude contexts.

for weakening the Equivalence Principle within classical logic. I think it's fair to say that each of these classical sub-options has a high cost: the weakenings of the Equivalence Principle are highly unnatural, and it isn't obvious that they wouldn't cripple the ordinary use of the notion of truth. That's a question that requires a more serious discussion than I can give here, but in any case, there is motivation for exploring the second choice, that of restricting classical logic.

More fully, the approach that Yablo and I share involves the use of a weakened logic—in particular, one without the law of excluded middle[2]—as one's general logic. This is compatible with allowing that the principle of excluded middle is correct (as a non-logical principle) in domains where peculiar concepts like 'true' are not involved or can be shown to be eliminable: e.g. no restriction on the use of classical logic within mathematics or within physics is required. The general logic that applies even when peculiar concepts like 'true' are in play cannot be intuitionist: intuitionism is inconsistent with the Equivalence Principle, just as classical logic is. And certain principles that are not valid in intuitionism, such as the equivalence between $\neg\neg A$ and A and between $\neg(A \wedge B)$ and $\neg A \vee \neg B$, can (and presumably should) be retained.

The idea of using logics of this sort to evade the paradoxes is not a new one, though early attempts turned out not to be consistent with the Equivalence Principle. The first work that can be viewed as demonstrating the consistency of the Equivalence Principle in a weakening of classical logic is Kripke [7]: in particular, the part of the paper that discusses the Kleene truth tables.[3] Kripke's discussion is entirely at the level of semantics, and there is more than one way to read a theory of truth off the semantics; but on what is I think the most natural way of reading a theory off the semantics (viz., one that takes the theory to be

[2] There is an alternative non-classical route that keeps excluded middle, but gives up instead the prohibition against accepting contradictions and the principle that contradictions imply everything. I won't be discussing that here.

[3] Kripke discusses both the weak and the strong Kleene truth tables; the strong are of more interest, and I'll confine my discussion to them.

the set of sentences in a particular fixed point) we get a theory that obeys the Equivalence Principle, in a logic without excluded middle.

Unfortunately, the logic that Kripke reconciles with the Equivalence Principle is so weak as to be very difficult to reason with: in particular, it does not contain a reasonable conditional. (We could define $A \rightarrow B$ as $\neg A \vee B$; but then the absence of excluded middle would preclude such things as $A \rightarrow A$ and $A \wedge B \rightarrow A$ from being valid. The lack of the former means that we also don't validate either half of the Tarski schema, i.e. either $True(\langle A \rangle) \rightarrow A$ or its converse.) The logic is expressively weak in other ways too; for instance, it has no way of registering the fact that Liar sentences are in some sense "pathological".

What Yablo and I are both interested in is showing the consistency of the Equivalence Principle in expressively richer logics, which contain at least a reasonable conditional (and in the case of my own account at least, contain ways of asserting pathology). Unfortunately the simplest ways of adding a new conditional to the logic lead to paradoxes of their own: it is easy to see that the conditional can't be a 3-valued truth function, and even the continuum-valued truth function of "fuzzy logic" can be shown to give rise to paradox ([11], [6]). Something more elaborate is required.

2 A simplified version of Yablo's account.

Yablo develops his account[4] using a 4-valued semantics, rather than the 3-valued semantics I used in [1]. I'm going to simplify things by using a 3-valued analog. All the virtues of his account that he cites survive this simplification; and the complaints I'll have about his account would arise for the 4-valued version as

[4]By "Yablo's account" I shall mean what he calls *Kripke-style possible world semantics*. Prior to giving this account he sketches a different account, which he calls *Field-style possible world semantics*, but this isn't one that he has any great interest in, and I don't either.

well. Besides simplification, the use of the 3-valued semantics leads to a more powerful logic for the connectives other than the conditional: with a 4-valued semantics one must apparently sacrifice either the principle that contradictions imply everything or the rule of disjunction elimination, whereas these can be jointly maintained on a 3-valued semantics.

2.1 Kripkean background.

Let L be a standard first order language, adequate to arithmetic and the theory of finite sequences (and hence adequate to the syntax of first order languages, even ones with uncountably many symbols), and let M be a classical model for L which is standard with respect to arithmetic and the theory of finite sequences. For notational convenience (i.e., to remove the need to talk of assignments of objects to free variables), I will assume that L has a name for every object in the domain of M (which if M is uncountable will mean that L contains uncountably many expressions). We can think of M as consisting of a domain, an assignment of an object in the domain to each name of L and of an operation on M to each function symbol of L, and an assignment of exactly one of the values 1 and 0 to each atomic sentence of L. We want to extend L to a language L^+ by adding a new 1-place predicate 'True' and a new 2-place conditional '$\rightarrow$'; when the syntax of L^+ is developed within L^+, 'True' will mean 'is a true sentence of L^+'. We also want to extend M to a 3-valued model M^+; M^+ will have the same domain as M, and treat the names and function symbols just as M does, and assign the same values to the atomic sentences of L that M assigns them; but it will also assign a "3-valued extension" to 'True', that is, it will assign one of the three values 1, $\frac{1}{2}$ and 0 to atomic sentences containing 'True'. If we denote this assignment to 'True' by T, then M^+ is given by the pair $\langle M, T\rangle$.

Once T is specified (in addition to M), that will be enough to determine a semantic value for every sentence of L^+ *not containing the* $\rightarrow$, by the following "strong Kleene" rules:

$$|\neg A|_{M,T} = 1 - |A|_{M,T}$$
$$|A \vee B|_{M,T} = \max\{|A|_{M,T}, |B|_{M,T}\}$$
$$|A \wedge B|_{M,T} = \min\{|A|_{M,T}, |B|_{M,T}\}$$
$$|\exists x A|_{M,T} = \max\{|A(x/t)|_{M,T}\}$$
$$|\forall x A|_{M,T} = \min\{|A(x/t)|_{M,T}\}$$

But the value of conditionals will require a separate rule, and the main question is going to be what this rule should be. For the moment we can bypass this crucial question in an uninformative manner, by introducing the idea of a *valuation for conditionals*. This is simply a function that assigns one of the three semantic values to each conditional in L^+ (including conditionals that have other conditionals as subformulas). Then if v is a valuation for conditionals, clearly the semantic value of every sentence of L^+ is determined by M^+ *together with* v. (The rules for connectives like $\neg$ and $\vee$ and $\exists$ merely need to be extended to include sentences with conditionals; e.g., the rule for $\neg$ is now that $|\neg A|_{M,v,T} = 1 - |A|_{M,v,T}$.)

Let's turn to the question of what the rule for 'True' should be. As mentioned, we want to maintain the Equivalence Principle; given that the language doesn't contain quotation marks, propositional attitude constructions and the like, we can state this in the following unrestricted form:

> **The Equivalence Principle (EP):** For any sentence C of L^+ in which a sentence A of L^+ occurs, the result of substituting $True(\langle A \rangle)$ for one or more occurrences of A has the same semantic value as does C.

If L^+ *hadn't contained the new conditional*, a necessary and sufficient condition

for attaining (EP) would be that the assignment of a 3-valued extension to 'True' be a *Kripkean fixed point*; that is,

> **Kripke Fixed Point Requirement (KFP):** For any sentence A of L^+, $|True(\langle A\rangle)|_{M,v,T}$ must be the same as $|A|_{M,v,T}$. (And if T doesn't denote a sentence of L^+, $|True(t)|_{M,v,T}$ must be 0.)

(Of course the subscript v isn't really needed when the language doesn't contain the conditional.) But with the conditional in the language, this is no longer sufficient for (EP): we need in addition a requirement on the valuation of conditionals, viz.

> **Transparency Requirement (TR):** For any *conditional* C of L^+ in which a sentence A of L^+ occurs, the result of substituting $True(\langle A\rangle)$ for one or more occurrences of A has the same semantic value as does C.

(Again, the conditionals in question include ones with other conditionals as subformulas.) But the requirement (KFP) on the assignment to 'True' and the transparency requirement (TR) on the valuation of conditionals together suffice for the Equivalence Principle, as a simple inductive argument shows.

Moreover, Kripke showed that satisfying (KFP) is easy: although he didn't discuss languages with additional conditionals, his argument can be extended to show that for any valuation for conditionals v, there are a wide variety of *fixed points over* v, that is, 3-valued extensions of 'True' that satisfy (KFP). In particular there is a *minimal fixed point over* v, i.e. one which assigns the value 1 or 0 to a sentence only if every fixed point over v assigns the same value to that sentence.

Let's simplify the notation. I will throughout be concerned with a single ground model M, so there will be no need to keep including it in the subscripts.

Moreover, from now on I'll be concerned only with pairs $\langle v, T\rangle$ for which T is a Kripke fixed point over v. And given this, there is really no need to keep v in the notation, for it is determined by T: v is that function that assigns to any conditional $A \rightarrow B$ whatever value $True(\langle A \rightarrow B\rangle)$ gets in T. I'll call this v_T; so we'll be concerned with assignments T to 'True' each of which is a Kripke fixed point over a unique valuation v_T. Given all this, the notation $|A|_{M,v,T}$ simplifies to $|A|_T$.

So to repeat, if v is any transparent valuation of conditionals, there are Kripke fixed points over v, and by evaluating sentences in accord with any such fixed point, the Equivalence Principle is guaranteed. And the evaluation agrees with the originally-given valuation M for sentences without 'True' or '$\rightarrow$ '.

2.2 The Yablo conditional.

The material in Section 2.1 (which is common ground between Yablo's approach and my approach in [1]) doesn't by itself provide much of a theory, since so far there is no serious account of how the conditional works: the valuation of conditionals v has been left completely arbitrary, except for the requirement that it be transparent. There is no guarantee that sentences that ought to come out logical truths (e.g. instances of $A \rightarrow A$ or $A \wedge B \rightarrow A$) will get value 1; indeed, every conditional could get value $\frac{1}{2}$, or every one could get value 0. Nor is there a guarantee that modus ponens will be validated. What we need is a reasonable way to assign values to conditionals.

Yablo's proposal—again, modified to fit a 3-valued framework—is elegantly simple. For any transparent valuation v (of conditionals), let $S[v]$ be the set of transparent valuations that extend v (i.e. which assign 1 to every conditional to which v assigns 1 and 0 to every conditional to which v assigns 0); since v is transparent, $v \in S[v]$, so $S[v]$ is not empty. For any nonempty set

S of valuations (of conditionals), let $w[S]$ be the valuation given as follows: for any A and B,

$w[S](A \rightarrow B)$ is

1 if $(\forall v)(\forall T)$(if $v \in S$ and T is a Kripke fixed point over v then $|A|_T \leq |B|_T$),

0 if $(\forall v)(\forall T)$(if $v \in S$ and T is a Kripke fixed point over v then $|A|_T > |B|_T$),

$\frac{1}{2}$ otherwise.

Clearly $w[S]$ is transparent if all members of S are. Moreover if v_1 extends v_2 then $S[v_1] \subseteq S[v_2]$; and if $S_1 \subseteq S_2$ then $w[S_1]$ extends $w[S_2]$; consequently, if v_1 extends v_2 then $w[S[v_1]]$ extends $w[S[v_2]]$.

Let a *Yablo fixed point* be a valuation v for which v is identical to $w[S[v]]$; that is, a valuation v such that for any A and B,

(YFP) $v(A \rightarrow B)$ is

1 if $(\forall u)(\forall T)$(if u is a transparent extension of v and T is a Kripke fixed point over u then $|A|_T \leq |B|_T$)

0 if $(\forall u)(\forall T)$(if u is a transparent extension of v and T is a Kripke fixed point over u then $|A|_T > |B|_T$),

$\frac{1}{2}$ otherwise.

Yablo thinks that a reasonable valuation of conditionals should be a Yablo fixed point. He says that this gives the conditional an appealing modal-like semantics: if v is a Yablo fixed point and T is a Kripke fixed point over it, then the "possible worlds" accessible from $\langle M, v, T\rangle$ are the $\langle M, u, U\rangle$ for which u is a transparent extension of v and U is a fixed point over u. (Presumably the idea is that the "worlds" in the semantics don't really represent *possibilities* in any normal sense—those are fixed by the ground world M—but rather *minimally adequate ways of valuating sentences*. A conditional sentence $A \rightarrow B$ is "modal" in

something like the way that a sentence DA asserting what is determinately the case is modal in many treatments of vagueness: the value of $A \rightarrow B$ or DA is determined by looking at *a range of* valuations of the sentences A and B to which it applies, rather than just at a single valuation.)

It is easy to show that there are Yablo fixed points; indeed, there is a natural Kripke-like construction of the smallest one. We define a transfinite sequence of transparent valuations, as follows:

v_0 assigns each conditional the value $\frac{1}{2}$;

$v_{\alpha+1}$ is $w[S[v]]$;

v_λ is the minimal extension of each v_β for $\beta < \lambda$, when λ is a limit ordinal.

[Or to write the successor clause without the abbreviations: $v_{\alpha+1}(A \rightarrow B)$ is

1 if $(\forall u)(\forall T)$(if u is a transparent extension of v_α and T is a Kripke fixed point over u then $|A|_T \leq |B|_T$)

0 if $(\forall u)(\forall T)$(if u is a transparent extension of v_α and T is a Kripke fixed point over u then $|A|_T > |B|_T$),

$\frac{1}{2}$ otherwise.]

A trivial inductive argument shows that if $\alpha < \beta$ then v_β extends v_α; so by a standard fixed point argument analogous to Kripke's, there must be an ordinal ξ such that $(\forall \alpha \geq \xi)(v_\alpha = v_\xi)$. v_ξ is *the minimal Yablo fixed point*.

Yablo's proposal (modified to fit 3-valued semantics) is to use this minimal Yablo fixed point v_ξ as the valuation for conditionals, and the minimal Kripke fixed point over v_ξ (call it Z_ξ) as the assignment to the truth predicate.

2.3 Discussion.

Aside from its elegant simplicity, Yablo's proposal has many attractive features. First, it clearly validates many desirable laws (i.e., gives all instances of them

value 1, in the case of sentences; preserves value 1, in case of inferences). For a *very* simple illustration, consider $A \wedge B \rightarrow A$; for any fixed point T over any valuation at all, $|A \wedge B|_T \leq |A|_T$, so the construction gives this conditional the value 1 at v_1 and hence at every valuation thereafter. Somewhat more interesting is the inference from $(A \rightarrow B) \wedge (A \rightarrow C)$ to $A \rightarrow (B \wedge C)$. If the premise gets value 1 at v_ξ then by the Yablo fixed point property (YFP), $|A|_T \leq |B|_T$ whenever v_T is a transparent extension of v_ξ, and also $|A|_T \leq |C|_T$ in the same circumstances; from which it follows that $|A|_T \leq |B \wedge C|_T$ in those circumstances, and hence that $A \rightarrow (B \wedge C)$ has value 1 at v_ξ. Modus ponens is validated as well: if $A \rightarrow B$ has value 1 at v_ξ, then for any fixed point T over a transparent extension of v_ξ, $|A|_T \leq |B|_T$; so in particular when Z_ξ is the minimal Kripke fixed point over v_ξ, $|A|_{Z_\xi} \leq |B|_{Z_\xi}$; so if A has value 1 at Z_ξ, so does B. A list of some other laws that are validated in the theory is given in [12]. (The list is of things validated in the 4-valued semantics; but anything validated in the 4-valued is validated in the 3-valued as well.)

Second, the quantification over non-minimal Kripke fixed points (in the definition of the w operator and hence in (YFP)) produces very intuitive results for conditionals in which the antecedent and consequent could consistently be assigned values in more than one way. (These cases form the basis of one of Yablo's objections to my account in [1], the objection from "insufficient strictness".)[5]

Consider for instance the conditionals $L \rightarrow I$ and $I \rightarrow L$, where L is a Liar sentence (one which asserts its own untruth) and I is a Truth-Teller (one which asserts its own truth). On Yablo's account, L and I get value $\frac{1}{2}$ in the designated fixed point Z_ξ (the minimal fixed point over v_ξ); but whereas L gets value $\frac{1}{2}$ in all other fixed points as well, I gets value 0 in some fixed points over

[5] In my opening remarks I mentioned that Yablo gives three objections to my account in [1]. The other two won't be considered until Section 4.

v_ξ and value 1 in others. Because of this, the conditionals $L \rightarrow I$ and $I \rightarrow L$ get value $\frac{1}{2}$ on Yablo's account. I agree with Yablo that this seems intuitively right; my semantics in [1] (which also gave value $\frac{1}{2}$ to L and I) was unintuitive in giving these conditionals value 1. That's one illustration of why Yablo says, correctly I think, that my conditional in [1] *wasn't sufficiently strict*.

For a second example in which quantifying over non-minimal as well as minimal fixed points produces a desirable strictness, consider the conditional $I \rightarrow \neg I$ and its converse, where again I is a truth-teller. An oddity of my own account in [1] was that these got value 1. ($I \rightarrow I$ got value 1 as well, as it should.) In the case of a Liar sentence L, it is inevitable that $L \leftrightarrow \neg L$ as well as $L \leftrightarrow L$ should get value 1: that $L \leftrightarrow \neg L$ gets value 1 follows from the Equivalence Principle (and with excluded middle gone, it doesn't lead to contradiction). But it does seem that we ought to minimize the number of sentences A for which $A \leftrightarrow \neg A$ (as well as $A \leftrightarrow A$) holds, and it seems undesirable that it holds for truth-tellers. The quantification over non-minimal as well as minimal fixed points in Yablo's account is enough to guarantee that $I \rightarrow \neg I$ and its converse each get value $\frac{1}{2}$.

Yablo gives a third example along the same lines. Suppose Jones and Smith each say that what the other says is not true. Any account that respects the naive theory of truth (the Equivalence Principle plus the Tarski biconditionals) will yield $J \leftrightarrow \neg S$ (and so by the symmetry of the situation, any reasonable 3-valued account will give each of J and S the value $\frac{1}{2}$). But my account in [1] also yielded $J \leftrightarrow S$, which seems undesirable; and by quantifying over non-minimal fixed points as well as minimal ones, Yablo's account avoids this.[6]

It's worth noting that while these examples clearly illustrate the virtues

[6] For another illustration of the (3-valued) Yablo account, let L_1 and L_2 be two Liar sentences (each asserting its own untruth), and I_1 and I_2 two truth-tellers. The Yablo semantics yields that $L_1 \leftrightarrow L_2$ gets value 1, seemingly making the Liar sentence essentially unique, but yields that $I_1 \leftrightarrow I_2$ gets value $\frac{1}{2}$. (The reason for the 'seemingly' will appear in note 9.)

of quantifying over non-minimal fixed points in the clause for a conditional *having value* 1, they don't turn at all on the fact that Yablo also quantifies over non-minimal fixed points in the clause for a conditional *having value* 0. Here is a minor variant of Yablo's account: instead of using the operator w of Section 2.2, we use the following operator w^*, where Z_v stands for the minimal Kripke fixed point over v:

$w^*[S](A \rightarrow B)$ is

1 if $(\forall v)(\forall T)$(if $v \in S$ and T is a Kripke fixed point over v then $|A|_T \leq |B|_T$),

0 if $(\forall v)$(if $v \in S$ then $|A|_{Z_v} > |B|_{Z_v}$),

$\frac{1}{2}$ otherwise.

This leads to a *Yablo* fixed point*, by the same argument; that is, a v such that

(YFP*) $v(A \rightarrow B)$ is

1 if $(\forall u)(\forall T)$(if u is a transparent extension of v and T is a Kripke fixed point over u then $|A|_T \leq |B|_T$)

0 if $(\forall u)$(if u is a transparent extension of v then $|A|_{Z_u} > |B|_{Z_u}$),

$\frac{1}{2}$ otherwise.

This account strikes me as slightly more natural than the actual Yablo account, but the differences will not matter to anything that follows.[7] Both accounts have the virtues cited above.

Despite its virtues, I have reservations about Yablo's account (and the Yablo* variant of it). Many of these reservations center on what it says (or doesn't say) about conditionals that have other conditionals embedded within them.

In the first place, let me note a point that Yablo himself cites as a weakness in his account: that on the account there seem to be no significant

[7] An example of the difference: if I is a Truth-teller, $I \rightarrow 0 = 1$ gets value 0 on Yablo*, $\frac{1}{2}$ on Yablo. (Whereas if L is a Liar, $L \rightarrow 0 = 1$ gets value 0 on both.)

validities involving nested conditionals. Among the laws one might expect to hold of a conditional are the following:

$$(A \rightarrow \neg B) \rightarrow (B \rightarrow \neg A);$$

$$(A \rightarrow \neg A) \rightarrow \neg(\top \rightarrow A), \text{ where } \top \text{ is a tautology such as } B \rightarrow B;$$

and the inferences

$$A \rightarrow B \vdash (C \rightarrow A) \rightarrow (C \rightarrow B);$$

and $$A \rightarrow B \vdash (B \rightarrow C) \rightarrow (A \rightarrow C).$$

These all fail on Yablo's account and the Yablo* variant (though associated rules without embedded conditionals, like $A \rightarrow \neg B \vdash B \rightarrow \neg A$, hold).

The loss of the last pair of inferences is quite important: it blocks the proof of the substitutivity of equivalents (the inference from $A \leftrightarrow B$ to $X_A \leftrightarrow X_B$, where X_B is the result of substituting B for one or more occurrences of A in X_A). And indeed, that substitutivity principle fails dramatically in the Yablo (and Yablo*) semantics. For instance, though $A \leftrightarrow (A \vee A)$ is valid, $[A \rightarrow C] \leftrightarrow [(A \vee A) \rightarrow C]$ isn't. For consider any A and C for which $A \rightarrow C$ gets value $\frac{1}{2}$ in v_ξ; then $A \vee A \rightarrow C$ also gets value $\frac{1}{2}$ in v_ξ. So in some extensions of v_ξ, one of $A \rightarrow C$ and $A \vee A \rightarrow C$ will get value 0 while the other one gets a value different from 0, and this means that $[A \rightarrow C] \leftrightarrow [(A \vee A) \rightarrow C]$ will get value $\frac{1}{2}$. Clearly this situation arises because the extensions of v_ξ that one quantifies over in the truth conditions at v_ξ can be extraordinarily badly behaved; one might try to fix this by modifying the theory so that badly behaved valuations are excluded from the start, but it isn't at all evident how this might be done without destroying the proof that the valuation of conditionals reaches a fixed point.[8]

[8] The fixed point proof above relied on the fact that when S is a set of transparent valuations, $w[S]$ is transparent; but if S is, say, a set of (transparent) valuations that validate the rules $A \rightarrow B \vdash (C \rightarrow A) \rightarrow (C \rightarrow B)$ and $A \rightarrow B \vdash (B \rightarrow C) \rightarrow (A \rightarrow C)$, $w[S]$ need not validate those rules. (If $A \rightarrow B$ doesn't get value 1 throughout S then we can have $|C \rightarrow A| > |C \rightarrow B|$ for some members of S. This can happen even if $|A| \leq |B|$ throughout S; in that case, $A \rightarrow B$ gets value 1 in $w[S]$ but $(C \rightarrow A) \rightarrow (C \rightarrow B)$ doesn't.)

The failure of substitutivity means that though both the Tarski biconditionals and the Equivalence Principle hold in the Yablo semantics, one can't infer the latter from the former in the way one might have expected.[9]

A related point is that Yablo's claims to have given the $\rightarrow$ a modal semantics seem considerably overstated, for when a conditional is embedded inside another conditional, the "modal semantics" applies only to the outer conditional, not to the occurrence of the conditional embedded inside it. The analogs of "possible worlds" on the Yablo semantics are the Kripke fixed points over extensions of the Yablo fixed point v_ξ; in an embedded conditional, the inner conditionals are thus evaluated at extensions of v_ξ. But these extensions of v_ξ are, for the most part, not themselves Yablo fixed points, and the evaluation of conditionals at them does not proceed by considering the values of their antecedents and consequents at other worlds. Rather, the evaluation of conditionals at them is just built in by brute force, it's built into the specification of which extension of v_ξ is in question.

It might be thought that some of the issues I've raised could be resolved by iterating Yablo's construction. To explain: another way of looking at the Yablo construction is as starting from the set S_0 of all transparent valuations, and for each α letting $S_{\alpha+1}$ be $S[w[S_\alpha]]$, taking intersections at limits; this has a nonempty intersection S_∞, and v_ξ is $w[S_\infty]$. On this way of looking at things, what we've done in constructing the fixed point is to successively throw out valuations that aren't candidates for our final valuation, and construct the next valuation in the sequence by quantifying only over what remains. But from this viewpoint, it would seem we could go further: the only real candidates for our final valuation are Yablo fixed points; now that we know there are some, why not introduce a new sequence of sets starting with the set S_0^1 of all Yablo fixed

[9] It also means that one can't infer from the fact that the biconditional connecting two Liar sentences gets value 1 that one Liar sentence is substitutable for another in all contexts; and in fact it isn't, as reflection on the example in the previous paragraph should make clear.

points, with its corresponding valuation $v_0^1 = w[S_0^1]$, and successively decrease the former and build up the latter until a new "second level Yablo fixed point" is reached? (And we might then want to iterate still further.) The motivation for the Yablo account would seem to extend to this iterated version, and the iterated version would seem as if it might give rise to a more fully modal semantics.

Unfortunately, the iterated version breaks down right at the start. To see this, consider the Curry sentence K, which says $True(\langle K \rangle) \rightarrow 0 = 1$. This has value $\frac{1}{2}$ not only in v_ξ but in every ("first level") Yablo fixed point.[10] So defining S_0^1 and v_0^1 as above, every member of S_0^1 gives K the value $\frac{1}{2}$; so by the valuation rules for the conditional, v_0^1 gives K the value 0, and hence is not even a first level Yablo fixed point, i.e. not a member of S_0^1. As a result, the iterated procedure does not evolve toward a fixed point, and thus a fundamental feature of Yablo's account would be destroyed.[11]

Even if iteration is a bad idea, one might still hope that the (uniterated) Yablo account could be given a genuine modal semantics, by finding some accessibility relation on the set $S[v_\xi]$ (the set of Kripke fixed points over transparent extensions of v_ξ) such that for each T in $S[v_\xi]$,

$v_T(A \rightarrow B)$ is

1 if $(\forall U \in S[v_\xi])$(if U is accessible from T then $|A|_U \leq |B|_U$)

0 if $(\forall U \in S[v_\xi])$(if U is accessible from T then $|A|_U > |B|_U$), and there are U accessible from T

$\frac{1}{2}$ otherwise.

[10] Proof: If K has value 1 at a valuation v, then it has value 1 in every extension of v and so in every Kripke fixed point over such an extension; but then the evaluation rules for $True(\langle K \rangle) \rightarrow 0 = 1$ yield that it has value 0 at $w[S[v]]$, so v can't be a Yablo fixed point. Similarly, if K has value 0 at a valuation v, then it has value 0 in every extension of v and so in every Kripke fixed point over such an extension; but then the evaluation rules for $True(\langle K \rangle) \rightarrow 0 = 1$ yield that it has value 1 at $w[S[v]]$, so v can't be a Yablo fixed point. (This works for the Yablo* rules as well.)

[11] As we'll see later, there are ways to give up the demand for fixed points, and perhaps the suggestion of iterating the Yablo construction could be pursued in that context. But I think that the synthesis to be suggested in Section 3 is simpler.

(Or this with the additional restriction in the 0 clause that U be minimal over v_U, if you prefer the Yablo* approach.) I don't think this is at all promising. First, note that for every T, there must be U accessible from it: otherwise every conditional would have value 1 at T, which can't happen since $0 = 0 \rightarrow 0 = 1$ clearly gets value 0 in every member of $S[v_\xi]$. (So the second conjunct of the clause for value 0 can be dropped.) Now consider the Curry sentence K (which is equivalent to $K \rightarrow 0 = 1$); using that equivalence, the above yields

$v_T(K)$ is

1 if $(\forall U \in S[v_\xi])$(if U is accessible from T then $|K|_U = 0$)

0 if $(\forall U \in S[v_\xi])$(if U is accessible from T then $|K|_U > 0$)

$\frac{1}{2}$ otherwise.

And since K is a conditional, $|K|_U$ is just $v_U(K)$. We know that there are in $S[v_\xi]$ plenty of valuations where K has value 1; so in all nodes accessible from such a node (and there are some accessible from it), K has value 0. Clearly then the accessibility relation can't be reflexive, and can't be connected in any nice way with the extendability relation. It also can't be transitive. For consider a node s_2 where K has value 0 that is accessible from a node s_1 where K has value 1. Since K has value 0 at s_2, it must be that in all nodes s_3 accessible from s_2 (and there are some), K has value > 0; so such nodes s_3 can't be accessible from s_1 given that K has value 1 at s_1. I suspect that one could prove that there is no accessibility relation whatever that would work, but the above makes pretty clear that at the very least any one that did work would have to be extraordinarily unnatural.

The lack of a specifically modal semantics isn't particularly troubling, but what does seem to me *a bit* troubling is that the Yablo account comes with no compositional semantics at all: no model-theoretic semantics in which we can assign fine-grained semantic values to sentences in each model, in such a way

that semantic values of sentences built from them can be determined.[12] (The reason for the 'fine-grained' is that we *obviously* can't expect a semantics which is compositional with respect to the assignment of the *coarse-grained* values 0, $\frac{1}{2}$ and 1 to the sentence: we already know that the conditional is not a three-valued truth function.) In a 3-valued modal semantics, you can take the fine-grained value of a sentence to consist of the set of worlds at which it has value 1 together with the set of worlds at which it has value 0; using the accessibility relation, fine-grained values for complex sentences are determined from fine-grained values for simple ones. There are other, non-modal, ways in which one might specify fine-grained semantic values that behave compositionally.[13] But the blindness toward the properties of embedded conditionals on Yablo's account appears to rule out any significant compositional semantics.

I will conclude this section with a worry of a different sort about Yablo's account: that at least as it stands, its expressive power is too limited. One thing we should want in a treatment of the paradoxes is to be able to consistently express in the language the idea that certain sentences of the language (Liar sentences, Curry sentences, and Truth-teller sentences) are in some sense "pathological". But just as the addition of a conditional to the language must be done with extreme care, to prevent new arguments for inconsistency from arising, so too the addition of predicates like 'pathological' must be done with extreme care: we have to make sure that apparently paradoxical sentences like 'I am either untrue or pathological' don't actually lead to inconsistency with

[12] Qualification: Obviously there are trivial ways of getting "fine-grained values" that meet this condition; e.g., we could take the "fine-grained value" of A to be simply the function that assigns one of the values 0, $\frac{1}{2}$ and 1 to each sentence containing A! What we want, I assume, is a *non-trivial* compositional semantics, but I admit I don't know how to make the non-triviality requirement precise.

[13] Indeed, the approach of mine that Yablo is criticizing does have a fine-grained compositional semantics—two of them really—though they didn't appear in the paper that Yablo was primarily addressing. An algebraic semantics for it is spelled out in [3] and [4], and a broadly modal semantics, though one employing a richer structure than simply an accessibility relation, is spelled out in [2].

the Equivalence Principle. One way of trying to do this is by defining "pathology predicates" from the conditional; if one can do this, then the consistency of one's treatment of the conditional guarantees the consistency of one's pathology predicates. That is the course I took in [1]. If one's treatment of the conditional doesn't allow for this, then one needs to expand the language to include the appropriate pathology predicates before one has an adequate overall theory. My worry about Yablo's account is that his conditional does not appear to allow for the definition of adequate pathology predicates, so that until we know how they might be added non-definitionally, his account is incomplete.

I won't attempt to prove that there is no way to define adequate pathology predicates on his account, but the prospects don't look good. The usual way to define them is to first define an operator D meaning 'determinately', with the property that if A has value 1 or 0, DA has the same value, and if A has value $\frac{1}{2}$, DA must have value no greater than $\frac{1}{2}$. Also, we'd like it to be the case that if A has value $\frac{1}{2}$ then either DA, or DDA, or $DDDA$, or some further (possibly transfinite) iteration D^α of D applied to A must have value 0; there are some technical issues that block a complete implementation of this (having to do with the impossibility of a single fully general method of defining the set of transfinite iterations of D),[14] but we'd certainly hope that for all but the most *recherche* sentences this holds unproblematically. If such a determinately operator can be defined, then the pathology predicates will have the form $\neg D^\alpha True(x) \wedge \neg D^\alpha True(neg(x))$, where '$neg$' stands for negation.

[14] One needs the truth predicate to define D^λ for limit λ, and the required definitions get more complicated as λ gets more complicated. Because of this last fact, any precise definition of a set of predicates D^α extends only through a proper initial segment of the recursive ordinals, and a different method of defining the allowed iterations would have allowed for iterations through a larger initial segment. On a given method of defining the class of iterations, there are bound to be sentences A that get value $\frac{1}{2}$ for which no iteration D^α that is definable by that method is such that $D^\alpha A$ has value 0. What we seem to want is that whenever a sentence A gets value $\frac{1}{2}$, "there should be some method of defining iterations such that on some iteration definable by that method, $D^\alpha A$ gets value 0", but it is doubtful that precise sense can be made of this.

But how on Yablo's theory are we to define D? As remarked above, the fact that the conditional has a modal element, where the modality isn't a quantification over alternative possible worlds but rather over alternative ways of evaluating sentences in the actual world, does make for an intuitive connection between the conditional and the notion of determinateness. However, the ways of defining D that tend to work for other versions of the conditional are $\top \rightarrow A$ and $\neg(A \rightarrow \neg A)$ and minor variations of these, but these don't work on Yablo's: for instance, applying any iteration of them to a Truth-teller sentence or its negation leaves the value $\frac{1}{2}$, so $\neg D^{\alpha}True(\langle I\rangle) \wedge \neg D^{\alpha}True(\langle \neg I\rangle)$ is never 1, so we couldn't assert that truth-tellers have any degree of pathology.[15] Again, this is not a criticism of Yablo's account if that is viewed merely as a treatment of the conditional; I'm simply saying (i) that it isn't obvious that the theory can be attractively supplemented to contain pathology predicates, (ii) that unless it can, we don't have an adequate treatment of the paradoxes, and (iii) that it is worth looking at alternative conditionals from which the pathology predicates can be defined, since they don't raise this worry.[16]

Nothing I've said is intended to be anything close to a knock-down objection to Yablo's approach. My points are intended only as reasons why we might hope to do better, while preserving the good points of his approach. It's to this that I now turn.

3 The Basic Synthesis.

I think that the worrisome features of Yablo's account stem from the fact that even at the final stage (the minimal Yablo fixed point), the evaluation of con-

[15] This doesn't seem to depend at all on the inability to extend the iteration far enough that was discussed in note 14.

[16] I should point out that at the end of [12] Yablo expresses some skepticism about whether my definitions of pathology predicates in [1] are really adequate. If he's right, that tends to undermine this criticism of his own account. I don't think he is right, but the issue is too big to discuss here.

ditionals is done by quantifying over a large array of Kripke fixed points *over quite bad valuations for conditionals*: valuations which, though they extend the minimal Yablo fixed point, can do so in quite arbitrary and unprincipled ways.

I propose that we modify Yablo's—or rather, Yablo*'s—inductive sequence of v_αs, so as to avoid this. The central change is in the rule for successors: instead of the Yablo* rule that $v_{\alpha+1}(A \rightarrow B)$ is

1 if $(\forall u)(\forall T)$(if u is a transparent extension of v_α and T is a Kripke fixed point over u then $|A|_T \leq |B|_T$),

0 if (∀u)(if u is a transparent extension of v_α then $|A|_{Z_u} > |B|_{Z_u}$),

$\frac{1}{2}$ otherwise,

I propose that we omit the quantification over extensions of v_α, and take $v_{\alpha+1}(A \rightarrow B)$ to be

1 if $(\forall T)$(if T is a Kripke fixed point over v_α then $|A|_T \leq |B|_T$),

0 if $|A|_{Z_{v_\alpha}} > |B|_{Z_{v_\alpha}}$,

$\frac{1}{2}$ otherwise.

This looks like a simple revision, but in fact it has a drastic consequence for the mathematical character of the theory: the rule is no longer monotonic, that is, $v_{\alpha+1}$ will no longer be an extension of v_α. Because of this, a change at the limit stage is called for: the appropriate rule is now that $v_\lambda(A \rightarrow B)$ is

1 if $(\exists \beta < \lambda)(\forall \gamma)(\forall T)$(if $\beta \leq \gamma < \lambda$ and T is a Kripke fixed point over v_γ then $|A|_T \leq |B|_T$),

0 if $(\exists \beta < \lambda)(\forall \gamma)$(if $\beta \leq \gamma < \lambda$ then $|A|_{Z_{v_\gamma}} > |B|_{Z_{v_\gamma}}$),

$\frac{1}{2}$ otherwise.

For the moment I'll keep v_0 as in the Yablo and Yablo* accounts—assigning $\frac{1}{2}$ to every conditional—though I'll reconsider this in Section 5.

The fact that this rule is not monotonic means that the construction no longer reaches a fixed point for conditionals (i.e. an analog of the Yablo fixed

points). But two things can be said to ameliorate this. The more minimal is that there is nonetheless a natural sense in which the theory produces *ultimate values*: we can take the ultimate value of a sentence A to be

1 if $(\exists\beta)(\forall\gamma)(\forall T)$(if $\gamma \geq \beta$ and T is a Kripke fixed point over v_γ then $|A|_T = 1$),

0 if $(\exists\beta)(\forall\gamma)(\forall T)$(if $\gamma \geq \beta$ and T is a Kripke fixed point over v_γ then $|A|_T = 0$)

$\frac{1}{2}$ otherwise.

(Or we could stick to minimal Kripke fixed points in the clauses for 1 and/or 0 : it would make no difference in this context.) It is easily shown (using an analog of the Continuity Lemma of [1]) that for conditional sentences this is equivalent to

1 if $(\exists\beta)(\forall\gamma)(\forall T)$(if $\gamma \geq \beta$ and T is a Kripke fixed point over v_γ then $|A|_T \leq |B|_T$),

0 if $(\exists\beta)(\forall\gamma)$(if $\gamma \geq \beta$ then $|A|_{Z_{v_\gamma}} > |B|_{Z_{v_\gamma}}$),

$\frac{1}{2}$ otherwise;

this makes the ultimate value of a conditional analogous to its value at limits. The more substantial point, which plays an important role in ensuring that the logic works neatly, is that it can be shown that the construction eventually cycles and that there are certain special ordinals (*acceptable ordinals*) in the cycles at which every conditional gets its ultimate value (so that at minimal Kripke fixed points.over acceptable ordinals, every sentence gets its ultimate value). These claims were established at length in [1] for a somewhat simpler theory which was like this one except that the construction involved only the minimal Kripke fixed points; an inspection of the proof offered there shows that it carries over to the modified theory without any hitch.

This account gives rise to a richer set of laws than does Yablo's (or

Yablo*'s), especially with regard to embedded conditionals. And because of the quantification over non-minimal fixed points in the clause for conditionals having value 1, this account agrees with Yablo's on the cases cited early in 2.3 that my original account got intuitively wrong.

There are however some other examples of Yablo's where this account runs afoul of Yablo's intuitions. These examples were part of the basis for two of Yablo's criticisms of my earlier theory that I haven't yet discussed.

One of the examples is the Conditional Truth-teller. Yablo gives two slightly different versions of this, and the more complex version raises issues I'll defer till Section 5; but in its simplest version, this is a sentence I^* that asserts that it is true if $0 = 0$; that is, I^* is equivalent to $0 = 0 \rightarrow True(\langle I^* \rangle)$. My account in [1] gave this sentence the value 0, and the account just sketched does so as well; whereas Yablo thinks it ought to get value $\frac{1}{2}$, which is what his account delivers.[17] Yablo's reason for thinking, independent of his theory, that $\frac{1}{2}$ is the appropriate answer, seems to be as follows: any of the assignments 0, 1 and $\frac{1}{2}$ to the sentence seem consistent with obvious principles; and in situations where this is so, the value $\frac{1}{2}$ (which is best thought of not as a value on par with 0 and 1, but rather as the absence of the values 0 and 1) is the only non-arbitrary assignment to make. (He calls this the "arbitrariness objection" to my semantics.)

The reason that I resist this argument is that on the account of determinateness I'll give, I^* will be seen to, in effect, assert its own determinate truth. (This will be evident shortly; for now I'll just say that as in the case of Yablo's theory, the quantification over "nearby valuations" in the clauses for

[17] In the account I've sketched, it's easy to prove inductively that I^* gets value 0 in all v_α for $\alpha \geq 1$. (This relies on the fact that the starting valuation v_0 gave this conditional a value less than 1.) Why does I^* get value $\frac{1}{2}$ on Yablo's account? If it had another value, there would have to be a first ordinal at which it had that value, which is easily seen to be a successor $\beta + 1$; the value at β is $\frac{1}{2}$. But then there are transparent extensions of v_β in which I^* gets value 1, which is incompatible with $v_{\beta+1}(I^*)$ being 0 ; and the fact that I^* gets value less than 1 at v_β is itself incompatible with $v_{\beta+1}(I^*)$ being 1.

the conditional gives the conditional an intuitive connection with the notion of determinateness.) And given that I^* calls itself determinately true, it couldn't be true without being determinately true; in contrast, it can be false without being determinately false. So the situation with regard to I^* is not really symmetric: it is much easier for it to be false than to be true. Because of this, it is not particularly surprising if it ends up having value 0 (whereas it would seem surprising if it ended up having value 1). I don't say that this makes it pretheoretically obvious that the sentence should come out having value 0 rather than value $\frac{1}{2}$, but I do think it removes the appearance that the assignment of value 0 is arbitrary.

Yablo's other example (which he uses in his "groundedness objection") involves an infinite chain of sentences B_1, B_2, ... Each B_i asserts that if it is true, so is the next one: $True(\langle B_i \rangle) \rightarrow True(\langle B_{i+1} \rangle)$. My account in [1], and the account above, declares that each of these B_is is true; i.e. it gives value 1 to each claim $True(\langle B_i \rangle)$ and hence to each B_i. (This is independent of the decision to assign value $\frac{1}{2}$ to conditionals at the initial stage of the revision procedure.)[18] Yablo says that while this assignment of values to the B_is isn't *arbitrary* (there are reasons why we shouldn't view the B_is as false), it is objectionable because the assignment is *ungrounded*: "To suppose that B_i is true is to suppose it has a true antecedent. But then its truth is owing to the truth of its consequent $True(\langle B_{i+1} \rangle)$, with the buck being passed forever down the line." (p. 320)

[18] It is completely obvious that the three somewhat natural assignments at the initial stage—those that assign the same value to each B_i—yield the value 1 for each B_i at acceptable points: they reach this value at the very next stage, and it can never change after that. But in fact *any assignment of values whatever* yields the value 1 for each i at acceptable points (indeed, for all points from stage $\omega + 1$ on). For in the first place, no B_i can be assigned 0 at two successive stages. It follows from this (using the Continuity Lemma of [1], which carries over to the present theory and is independent of the valuation used at stage 0) that no B_i can be assigned 0 at stage ω (or any other limit stage). Also, if a B_i is assigned 1 at stage ω, the Continuity Lemma says that it is assigned 1 at all finite stages after stage n for some natural number n, and from this it follows that for all $j \geq i$, B_j has value 1 at stage n; which in turn implies that at all stages after $n + i$ (not just the finite stages), every B_j has value 1. The only alternative left to consider is that every B_i is assigned $\frac{1}{2}$ at stage ω; but in that case, every B_i gets value 1 at all later stages.

Yablo's own account gives each B_i the value $\frac{1}{2}$.[19]

I grant that this ungroundedness consideration has a certain pre-theoretic pull, but I think that there is at least equal weight to the following: because the sequence B_1, B_2, ... is isomorphic to the sequence with B_1 dropped, each B_i is "essentially equivalent" to the next. So in the conditionals $B_i \rightarrow B_{i+1}$ the antecedent is "essentially equivalent" to the consequent, and that motivates assigning the conditional the value 1. (From which it follows that each conditional $True(\langle B_i \rangle) \rightarrow True(\langle B_{i+1} \rangle)$ should get value 1, i.e. that each B_i should get value 1.)

Again, my claim isn't that this pre-theoretic argument is decisive; the point is only that while groundedness considerations give some intuitive support to assigning the value $\frac{1}{2}$ to the B_is, the alternative assignment of value 1 has some intuitive support as well. The intuitive considerations don't seem to me nearly strong enough to decide between two theories that yield different verdicts about the case.

4 Further Discussion.

The considerations that incline me to favor the account in Section 3 over Yablo's are

(I) its richer set of laws, including especially laws involving embedded conditionals;

(II) its being more amenable to assigning "fine-grained semantic values" that behave compositionally;

[19] Suppose not, and let α be the first stage in the Yablo construction at which at least one B_i gets value 0 or 1. α clearly isn't 0 or a limit, so it is of form $\beta + 1$. So v_β assigns each conditional B_i (or equivalently, each $B_i \rightarrow B_{i+1}$) the value $\frac{1}{2}$, and so for each i, v_β has a transparent extension that assigns B_i the value 1 and B_{i+1} the value 0. Since the value of B_i is greater than that of B_{i+1} in this extension, and no greater than it in v_β itself, the value of each $B_i \rightarrow B_{i+1}$ must be $\frac{1}{2}$ in $v_{\beta+1}$, contrary to supposition.

and (III) its giving rise to a natural determinately operator and hence to a sequence of pathology predicates.

I've already explained why I find Yablo's theory unsatisfying in these respects. I'll be brief in indicating how the account in Section 3 fares in these respects, because I've discussed these points elsewhere in connection with the account offered in [1], and the differences between that account and the account offered in Section 3 make little difference to the three points in question.

Let's begin with (III). The definition of the determinately operator offered in [1] was:

DA is $(\top \rightarrow A) \wedge A$, where $\top$ is some trivial truth.

Employing this definition in connection with the revision theory of the previous section, we get that in any fixed point U over $v_{\alpha+1}$, $|DA|_U$ is

1 if $(\forall T)$(if T is a Kripke fixed point over v_α then $|A|_T = 1$) and $|A|_U = 1$,

0 if $|A|_{Z_{v_\alpha}} < 1$ or $|A|_U = 0$,

$\frac{1}{2}$ otherwise.

And by the monotonicity property of Kripke fixed points, the quantification over non-minimal fixed points in the clause for value 1 is redundant: that clause is equivalent to

1 if $|A|_{Z_{v_\alpha}} = 1$ and $|A|_U = 1$.

As a result, the behavior of the determinately operator at successors is precisely the same as in [1]. And the same holds for limits, as is easily seen.[20]

[20] It is important that the theory in Section 3 took off from the Yablo* account rather than the Yablo account, i.e. that there was no quantification over non-minimal fixed points in the clause for a conditional having value 0; for a quantification over them in *that* clause would *not* be redundant for conditionals of form $\top \rightarrow A$. The resulting determinately operator would not only differ from that in [1], it would be inadequate: when I is a Truth-Teller, $|D^\alpha I|$ would never be 0, no matter how high the α.

That's the same criticism I made of the attempt to define a determinately operator in Yablo's own account, raising the question of whether the shift to the Yablo* variant would have evaded the criticism of his account. The answer is no: the problem for the Yablo account arises prior to the consideration of non-minimal fixed points, it arises already from the consideration of extensions of the base valuation.

Basically, then, the entire theory of the determinately operator, developed at some length in [1], carries over to the current theory. Let me just mention a few salient points. First, as mentioned in the discussion of Yablo, the operator gives rise to a series of "pathology predicates" $P_\alpha(x)$, each defined as $\neg D^\alpha True(x) \wedge \neg D^\alpha True(neg(x))$ where D^α is the α^{th} iteration of D (transfinite iterations being definable using the truth predicate, through a proper initial segment of the recursive ordinals, according to some fixed method). These have the following properties. (Here Δ is any acceptable ordinal, i.e. ordinal at which the values of conditionals coincide with their ultimate values.)

(i) They all have the same anti-extensions; that is, for any sentence A, and any α and β, $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = 0$ if and only if $|P_\beta(\langle A\rangle)|_{Z_\Delta} = 0$.

(ii) They have strictly increasing extensions; that is, if $\alpha < \beta$ (but β isn't so large that D^β and hence P^β are undefined) then

(a) for all A, if $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = 1$ then $|P_\beta(\langle A\rangle)|_{Z_\Delta} = 1$,

but (b) there are A for which $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = \frac{1}{2}$ and $|P_\beta(\langle A\rangle)|_{Z_\Delta} = 1$.

The simplest pathological sentences, such as Liar sentences, Curry sentences and Truth-Tellers, are all easily seen to be pathological at the first level—that is, the claim that they are P_1 has value 1. But use of the determinately operator or the pathology predicates can produce sentences for which one can't say whether they are P_α for small α, but can say that they are for high α. For instance, consider the "α^{th} level Hyper-Liar" L_α, which says $\neg D^\alpha True(\langle L_\alpha\rangle)$; then $|P_\beta(\langle L_\alpha\rangle)|_{Z_\Delta}$ is $\frac{1}{2}$ if $\beta \leq \alpha$, but 1 if $\beta > \alpha$. In [1] I've discussed a wide range of such transfinite sequences of paradoxical sentences, the members of each of which can all be declared pathological at some level.[21]

[21] As remarked in note 14, there is an inevitable arbitrariness in the system of transfinite iterations used in defining the D^αs and P^αs, which affects how far the iterations extend; no given method of defining a system of P_αs can be maximal. Because of this, on any given definition of the class of P_αs, one can't reasonably demand that for every sentence A of the language, either $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = 0$ for all P_α or there is a P_α such that $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = 1$; there are bound to be some sentences A for which $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = \frac{1}{2}$ for all α for which P_α has been defined. But (1) it does seem intuitively (though I don't know how to make this precise enough

Moving on to issue (II) about compositional semantics, I mentioned (in a footnote) that the account in [1] admits compositional semantics in either of two formats: algebraic and broadly modal. Both can be extended to cover the account in Section 3, but I'll discuss only the broadly modal version, since it is closer to the semantics that Yablo proposes for his own account. And in fact the broadly modal version offered in [2] barely needs extension at all: it was offered in a very general form (so as to apply to vagueness as well as to the paradoxes), and what I will do here is mostly just formulate the special case appropriate to the semantics of Section 3. But there will need to be a very small modification, due to the fact that I decided not to quantify over non-minimal fixed points in the 0 clause for conditionals.

In the broadly modal semantics for the account in Section 3, the "possible worlds" are the (non-minimal and minimal) fixed points over the *persistent valuations*, i.e. those valuations that appear over and over in the revision process. The "actual world" is the minimal fixed point over the valuation v_Δ that occurs at acceptable ordinals, i.e. the valuation that gives the ultimate values of all conditionals. It is a feature of this space of worlds that every valuation in it extends the valuation at the actual world, just as in Yablo's theory; but only very special extensions of v_Δ are in the space, which is part of what's responsible for the differences from Yablo's account.

The other difference from Yablo is that rather than giving a simple modal semantics based on an accessibility relation, I give a more complicated one based on the idea that $\rightarrow$ is in Lewis's phrase [9] a "variably strict conditional"

to prove it) that for any A for which this holds, a natural extension of the iteration procedure to larger ordinals is possible which would make the sentence "pathological with respect to a larger ordinal". (2) Even putting extensions of the iteration procedure aside, sentences A for which $|P_\alpha(\langle A\rangle)|_{Z_\Delta} = \frac{1}{2}$ for all α for which P_α has been defined must be extremely *recherche*: in particular, they must be such that the final cycle of values gives the sentence a sequence of 1's *that is at least as long as the first ordinal that outruns the iteration procedure*, followed by something other than a 1 (or analogously for 0 instead of 1). I doubt that one can produce such sentences except by building into them an explicit reference to the somewhat arbitrary system of ordinal notations used in the iteration procedure.

(though not one of quite Lewis's sort).

Picture the space of worlds as on a cylinder. For each persistent valuation v, the fixed points over v occur on a line L_v parallel to the axis of the cylinder. Moreover, the circular order of these lines L_v (say in the "clockwise" direction) corresponds to the order of the valuations in the cycle: starting with L_{v_Δ} (where Δ is acceptable), the next line is $L_{v_{\Delta+1}}$, then $L_{v_{\Delta+2}}$, and so on until you reach a sufficiently high ρ for $v_{\Delta+\rho}$ to be identical to v_Δ, at which point you are back around the cylinder at L_{v_Δ}.

Now given this picture, we can describe what Lewis calls a system of "spheres of similarity" around each "world" T in the space—though here they aren't spheres, but sections of the cylinder parallel to the axis. More precisely, a sphere of similarity for any fixed point on a line L_{v_β} will consist, for some line L_{v_α} distinct from L_{v_β}, of the set $S_{v_\alpha v_\beta}$ of fixed points that are *on a line that is strictly between L_{v_α} and L_{v_β} in the clockwise order, or on L_{v_α} itself.* (The latter disjunct is to ensure that none of the spheres of similarity are empty.) We can now describe **what it is for a conditional $A \to B$ to have value 1 at a "world" T**: it **is for there to be a sphere of similarity around that world such that at all worlds in that sphere, the value of A is less than or equal to that of B.**

If I had kept the quantification over all worlds in the 0 clause, the account of what it is for $A \to B$ to have value 0 at a world would be the same, except with 'greater than' instead of 'less than or equal to'. But since I didn't, I need an additional piece of structure: one point on each of the lines, corresponding to the minimal fixed point over that valuation, must be distinguished. (If you like you can think of there being a distinguished cross section of the cylinder, on which these distinguished worlds lie.) Then $A \to B$ **has value 0 at a world T iff there is a sphere of similarity around that**

world such that at all *distinguished* worlds in that sphere, the value of A is greater than that of B.

A slight oddity here is that points in L_{v_β} are not in any of their own spheres of similarity. However, the minimal fixed point on L_{v_Δ} has a distinguishing feature: any sentence A gets value 1 at $Z_{v_\Delta}$when and only when there is a sphere of similarity around it throughout which A gets that value, and similarly for 0. This means that *it would make no difference if we allowed Z_{v_Δ} to appear in its own spheres of similarity.* It is this special feature of Z_{v_Δ}that makes it natural to single it out as "the actual world".

We can now take the fine-grained semantic value of any sentence to be its "positive extension"—the set of worlds at which it has value 1—together with its "negative extension"—the set of worlds at which it has value 0. Fine-grained values of negations, disjunctions, quantifications etc. are determined in the usual way, and fine-grained values of conditionals are determined as in the boldfaced claims above. This is perhaps a more complicated modal semantics than one might have hoped for (and I grant that it could use a philosophical justification), but it is a genuinely compositional semantics in a way that Yablo's was not.

Turning finally to the issue (I) about laws, it is not hard to verify that almost all of the laws for the conditional that were established in [1] for the account there carry over to the modified account of Section 3: this can be established either by direct appeal to the account there, or (more perspicuously) by the semantics just sketched. The only one that fails here is the relatively unimportant

B4* $\quad \neg[(C \rightarrow A) \rightarrow (C \rightarrow B)] \vdash \neg[A \rightarrow B]$;

but its much more important contrapositive

B4 $\quad A \rightarrow B \vdash (C \rightarrow A) \rightarrow (C \rightarrow B)$

does hold, as does $A \rightarrow B \vdash (B \rightarrow C) \rightarrow (A \rightarrow C)$, with the consequence that the general substitutivity principle holds on this theory. (In addition, several other laws valid in the account of [1] but not noted there are valid here as well, e.g. $\neg(A \rightarrow B) \rightarrow (B \rightarrow A)$ and the rule $\neg(A \rightarrow B) \vdash A \vee \neg B$.[22])

5 The Starting Valuation.

In the course of his "arbitrariness objection" to my earlier account, Yablo points out an anomaly in that account which remains in the synthesis of section 3. The problem arises because of the choice of initial valuation v_0. I should note that the only requirement that must be imposed on v_0 for the basic theory of [1] to be derivable was that v_0 be transparent, and because of this, I employed a simple choice of transparent valuation as my starting point, the one that assigns each conditional the value $\frac{1}{2}$. But though this choice had no effect on the theory developed, and affects the ultimate values of only a few very special sentences, Yablo's point is that it does produce results that seem anomalous in some examples.

The example that Yablo gives to illustrate this point is a modified version of the Conditional Truth-Teller discussed in Section 4. Instead of an I^* that asserts of itself that if $0 = 0$ then $True(\langle I^* \rangle)$, Yablo considers an I^{**} that asserts of itself that if $B \rightarrow B$ then $True(\langle I^{**} \rangle)$, where B is any conditional that one chooses. $(B \rightarrow B) \rightarrow True(x)$ is of course equivalent to $0 = 0 \rightarrow True(x)$ in the theory, so one might think that there could be no real difference between I^* and I^{**}, but that thought involves a fallacy. Compare the predicates

D_1 The number of syllables in x is divisible by two

and

[22] In [2] I mistakenly asserted the stronger conditional form of this rule. In fact the conditional form is invalid even in the account of [1]: take A to be the Curry sentence and B to be its negation.

D_2 The number of syllables in x is divisible by the smallest prime;

these predicates are equivalent, but on any standard method for producing self-referential sentences, the sentences E_1 and E_2 that assert $D_1(\langle E_1\rangle)$ and $D_2(\langle E_2\rangle)$ respectively will have opposite truth values since D_2 has an odd number of additional syllables over D_1.

Even so, it does seem odd that I^{**} should get a different value from I^*. Another oddity is that the value it gets is 1, which in light of the fact that I^{**} in effect says of itself that it is determinately true seems especially hard to motivate. The reason it gets this value is that $B \rightarrow B$, though it has value 1 at each fixed point over any valuation *from* v_1 *on* in the construction, nonetheless has value $\frac{1}{2}$ at all fixed points over the chosen starting valuation v_0. I^{**}, as a conditional, also gets value $\frac{1}{2}$ at this stage; so since its antecedent and consequent have the same value at this stage it gets value 1 at the next stage, and this guarantees that it will get value 1 at each stage after this.

We can avoid this particular anomaly by beginning the construction from a more "regular" starting valuation. I'm currently undecided as to which starting valuation would be best to use: the choices I've thought of all seem a bit *ad hoc*. (But as I've said, the choice of the starting valuation plays little role in the overall theory, as long as it is transparent, and the ultimate value of "most" sentences is independent of the choice of transparent starting valuation.) One possibility I've contemplated is using as a starting valuation the minimal Yablo (or minimal Yablo*) fixed point; this would increase the extent to which the current account was a synthesis of the account in [1] with Yablo's. But even this wouldn't altogether avoid the sort of anomaly that Yablo has raised, because of certain conditionals B_0 that should get value 1 but don't in Yablo's minimal fixed point, e.g. the conditionals of form $(A \vee A \rightarrow C) \rightarrow (A \rightarrow C)$ considered earlier. There is no problem with such a B_0 itself: unlike on the

Yablo and Yablo* accounts, it gets the desired value 1 as its *ultimate* value on the synthesized account, whatever the starting valuation. But now consider an alternative conditional truth teller I^{***} that says that if B_0 then I^{***} is true; this will end up with value 1 on the proposed starting valuation, which seems rather analogous to the anomaly above, though for a more marginal sentence. It may be that there is no way to avoid all such anomalies involving conditional truth-tellers, without a more substantial alteration in the account; how serious a defect this would be, and whether there is a better approach that avoids such anomalies while preserving the advantages of the account, are questions that I leave for the reader.[23]

References

[1] Hartry Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32:139–177, 2003.

[2] Hartry Field. The semantic paradoxes and the paradoxes of vagueness. In JC Beall, editor, *Liars and Heaps*. Oxford University Press, 2003.

[3] Hartry Field. The consistency of the naive theory of properties. *Philosophical Quarterly*, 54, 2004.

[4] Hartry Field. Is the liar sentence both true and false? In JC Beall and Brad Armour-Garb, editors, *Deflationism and Paradox*. Oxford University Press, 2004.

[5] Harvey Friedman and Michael Sheard. An axiomatic approach to self-referential truth. *Annals of Pure and Applied Logic*, 33:1–21, 1987.

[23] I am grateful to Josh Schechter for suggesting several improvements.

[6] Petr Hajek, Jeff Paris, and John Shepherdson. The liar paradox and fuzzy logic. *The Journal of Symbolic Logic*, 65:339–346, 2000.

[7] Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[8] Stephen Leeds. Theories of reference and truth. *Erkenntnis*, 13:111–129, 1978.

[9] David Lewis. *Counterfactuals*. Harvard University Press, Cambridge, MA, 1973.

[10] W. V. O. Quine. *Philosophy of Logic*. Prentice-Hall, Englewood Cliffs, 1970.

[11] Greg Restall. Arithmetic and truth in Łukasiewicz's infinitely valued logic. *Logique et Analyse*, 139–140:303–312, 1992.

[12] Stephen Yablo. New grounds for naive truth theory. In JC Beall, editor, *Liars and Heaps*. Oxford University Press, 2003.