INDICATIVE CONDITIONALS, RESTRICTED QUANTIFICATION, AND NAIVE TRUTH

HARTRY FIELD

Abstract. This paper extends Kripke's theory of truth to a language with a variably strict conditional operator, of the kind that Stalnaker and others have used to represent ordinary indicative conditionals of English. It then shows how to combine this with a different and independently motivated conditional operator, to get a substantial logic of restricted quantification within naive truth theory.

1. Introduction

The "naive" notion of truth, according to which for each sentence S of our language, the claim that S is true is equivalent to S itself,[1] appears at first blush to be doomed by the Liar paradox and other related paradoxes. But only at first blush: one of the lessons that can be drawn from Kripke 1975 was that naivety in a theory of truth can be retained if one is willing to give up the hegemony of classical logic. There is little reason to doubt the correctness of classical logic as applied to our most serious discourse, e.g. our most serious physical theories. But the semantic paradoxes arise because truth talk gives rise to some anomalous applications (e.g. "viciously self-referential" ones), and it's rash to assume that classical logic continues to be appropriate to these applications. Maybe we should generalize logic in a way that allows these anomalies to be treated non-classically, while enforcing classicality in situations where anomalies can't arise. Kripke's paper, in particular the parts concerning logics based on Kleene valuation schemes, suggests the possibility of naive truth in this setting: in particular, one can have naive truth in a logic that restricts the general application of excluded middle, but which reduces to classical logic in contexts where the anomalies of truth cannot occur.

It isn't *immediately obvious* that the best response to the paradoxes is to abandon the hegemony of classical logic while retaining the hegemony of naive truth—*prima facie*, the reverse seems at least as attractive. But the costs of restricting naive truth turn out to be extraordinarily high,[2] and so the program of trying to keep it by restricting the scope of classical logic is one well worth pursuing. Kripke 1975 was the first substantial step.[3]

---

[1] I ignore ambiguities, indexical elements, etc., so as to be able to concentrate on sentence-*types*. There are subtleties about how best to extend the idea of naive truth to token utterances, but I will not be concerned with those issues here.

[2] See Field 2008, Part II, for a review.

[3] In Kripke's paper, and in the present paper too, we keep the classical *structural* rules for validity: (a) validity is transitive (in the general form given by the Cut Rule), and (b) valid inference is a relation between a *set* of premises and a conclusion (as opposed e.g. to a multi-set, where the number of occurrences of the premise matter, as in logics without structural contraction). The use of substructural logics is unnecessary.

Kripke's paper by itself shows the possibility of naive truth only for languages of very limited expressive power. The question arises as to how far his ideas can be generalized, and on this there has been some progress in recent years. In particular, there are now techniques for generalizing it to include certain kinds of conditionals (despite the threat of Curry-like paradoxes)[4]. But one kind of conditional operator that has not been treated in the literature on naive truth is "variably strict" conditional operators of the sort that have been discussed by Stalnaker 1968, Lewis 1974, Pollock 1976, Burgess 1981, and many others. The rough idea of such a conditional is that it is true at a world w if and only if at all worlds x *where its antecedent is true but that are othersise only minimally different from* w, its consequent is true. (There are different ways of spelling out this rough idea, depending mostly on the assumptions made about a relation of relative closeness of worlds; in this paper I'll adopt a framework, Burgess semantics, that is as neutral as possible about this.) Variably strict conditionals are clearly non-monotonic ('If A then C' doesn't imply 'If A and B then C'); from which it pretty much follows that they are non-transitive.[5] (They are also non-contraposable.) Their non-monotonicity and resulting non-transitivity make them significantly different from the sort of conditionals heretofore discussed in the naive truth literature. The early parts of the present paper provide a method (actually more than one) of extending Kripke's theory to cover languages with such a variably strict conditional—including in Section 6 the important case of languages that also have another conditional operator for restricted quantification.

Proponents of variably strict conditionals have divided over how extensive their application is. Some, e.g. Lewis, have taken a variably strict operator to model only "counterfactual" or "subjunctive" conditionals of ordinary language, and have held that "indicative conditionals" of ordinary language are represented by the familiar '$\supset$'. But it's well known that understanding ordinary indicatives in terms of '$\supset$' is *prima facie* counterintuitive—e.g. on that understanding, "If I run for President, I'll be elected" comes out true, since I'm resisting all pressure to run—and nowadays it's more common to think, with Stalnaker, that the variably strict conditional account is applicable to ordinary indicative conditionals as well as "counterfactuals". The first six sections of this paper are neutral on this issue.

But I favor the Stalnaker position, and this is relevant to an important application of the material in the early sections to the logic of restricted quantification, in Section 7. Restricted quantification poses a serious challenge to naive truth theory. In such a theory there are already difficulties with properly handling ordinary restricted quantifications like "Every true sentence in Jones' book appeared earlier in Smith's", but the difficulties become far greater when one tries to come up with a plausible account of how these interact with conditionals in a way that validates plausible laws such as "If all A are B and y is A then y is B" and "If everything is B then all A are B". I've addressed this challenge before (Field 2014), but in a rather *ad hoc* manner; an ultimate goal of this paper is to answer the challenge without *ad hoc*ness, by bringing in a more general logic of indicative conditionals.

There is also no need to restrict reasoning by cases, or to embrace dialetheism.

[4]See Restall 2007 for a discussion of such paradoxes and of the difficulties that a naive truth theory must overcome if it is to handle them.

[5]'If A and B then A' is clearly valid for them, and with it, transitivity would imply monotonicity.

2. Two-valued and three-valued worlds models for the language of indicative conditionals

Let L be a language whose logical primitives are '$\neg$', '$\wedge$', '$\forall$', '$=$', a unary necessity operator '$\Box$', and a binary conditional operator '$\rhd$'. An additional conditional for restricted quantification will be added in Sections 5 and 6. For the moment, let's suppose that L doesn't contain "paradox-prone" terms like 'True' that will require special treatment. '$\rhd$' is supposed to represent the indicative and/or counterfactual conditional of English and be a "variably strict" conditional in the general ballpark of Lewis, Stalnaker, Pollock and Burgess. Of these semantics, Burgess's is the most general (that is, the others can be obtained by adding restrictions to it),[6] and I will consider both it and a slight modification of it. Both versions of the Burgess semantics are initially based on "2-valued worlds models", which I'll now describe. (For simplicity I'll assume that L has no individual constants or function symbols; also, that its only variables are first order.)

A *2-valued worlds model* M for L consists of

(i): A non-empty set W of worlds, perhaps with a distinguished non-empty subset $NORM$ of "normal" worlds. (Nothing central to this paper depends on allowing non-normal worlds; I do so simply for added generality. The definition of validity will be in terms of the normal worlds only, but allowing for non-normal worlds may affect which conditionals can be true at normal worlds.)

(ii): For each $w \in W$, a subset W_w of W and a pre-order (reflexive and transitive relation) $\leq_w$ on W_w.[7] (Think of W_w as the set of worlds "accessible from" w, and '$x \leq_w y$' as meaning "the change from w to x is no greater than the change from w to y".)

(iii): For each $w \in W$, a non-empty set U_w (the universe of w). Let U be the union of the U_w.

(iv): For each $w \in W$ and k-place predicate p, a function p_w from U^k (the set of k-tuples of members of U) to $\{0, 1\}$. (The set of k-tuples that get assigned value 1 is the *extension of* p in the model.) We require that the function $=_w$ (associated with '$=$') assigns 1 to $< o, o >$ for each $o \in U$ and assigns 0 to all other pairs.

(W, $NORM$, etc. can all vary from one model to another, so we should really write W_M, $NORM_M$, $W_{M,w}$, $\leq_{M,w}$, $U_{M,w}$ and $p_{M,w}$.) Regarding (iv), we could if we like impose the ("actualist") requirement that p_w never assign value 1 to k-tuples not in ${U_w}^k$; it won't matter for what follows.[8]

Regarding (ii), we could if we like impose additional conditions on W_w and $\leq_w$ for each $w \in W$, or at least for each w in $NORM$. (The distinction of non-normal worlds from normal ones only matters if some such additional conditions apply

[6] Not every defensible model of conditionals can be fit into the Burgess framework (or the slight modification of it to be mentioned soon). I suspect that the basic ideas of this paper can be adapted to plausible alternative models, but will not attempt to prove this.

[7] An alternative convention is to take $\leq_w$ to be a pre-order on the full W, and subject to the constraint that if $y \in W_w$ and $x \leq_w y$ then $x \in W_w$.

[8] In the 3-valued context to be introduced shortly, we could introduce a more thorough actualism, in which the p_w never assign value 0 or 1 to such k-tuples; in effect this would make ${U_w}^k$ rather than the full U^k the domain of p_w. But again, this would make no difference to the issues I'm concerned with.

only to normal worlds.) Indeed, one such condition is almost universally regarded as appropriate for indicative and counterfactual conditionals (at least for worlds w in $NORM$):

Weak Centering: $w \in W_w$, and for any x in W_w, $w \leq_w x$

That Weak Centering holds at least for worlds in $NORM$ is required if Modus Ponens for $\triangleright$ is to be valid, on the account of validity soon to be given, which involves preservation of value 1 at normal worlds.[9] (Modus Ponens has been questioned for indicative conditionals (McGee 1985), but the grounds for doing so seem weak in the context of the semantics for variably-strict conditionals.)[10]

In addition to Weak Centering, Lewis, Stalnaker, Pollock and many others also accept one or more of the following conditions (for all worlds or just for normal ones):

Strong Centering: $w \in W_w$, and for any x in W_w other than w, $w <_w x$ (i.e. $w \leq_w x$ and not$(x \leq_w w)$)

No Incomparabilities: for any x, y in W_w, either $x \leq_w y$ or $y \leq_w x$

No Ties: for any distinct x, y in W_w, not both $x \leq_w y$ and $y \leq_w x$

Limit Condition: the relation $<_w$ is well-founded.

What follows will be completely neutral as to which if any such conditions are imposed, except for occasional reminders that restricting to models with Weak Centering (at least at normal worlds) is advantageous.[11]

To simplify the presentation of the semantics I adopt the usual trick of expanding the language to contain a new name for each object in U; call the expanded language L^+. (The expansion depends on the underlying model, so we should really write L^+_M.) I'll consider two ways of evaluating the sentences of L^+ in M.

---

[9]Demanding Weak Centering at non-normal worlds as well as normal ones would lead in addition, in the current 2-valued framework, to the validity of the inference from $C \triangleright A$ and $C \triangleright (A \triangleright B)$ to $C \triangleright B$. If we want Modus Ponens without getting that even for 2-valued sentences, we need the flexibility provided by non-normal worlds. In general, the point of non-normal worlds is to provide such added flexibility as to what comes out valid.

I've said that nothing in this paper depends on making use of such added flexibility: there will be no need to have the flexibility in the logic that includes 'True' if one doesn't utilize it in the base logic without 'True'. This may seem surprising: we presumably want Modus Ponens for $\triangleright$, but we don't want the law just cited since by taking A to be C we'll be led to the inference from $C \triangleright (C \triangleright B)$ to $C \triangleright B$, which in combination with Modus Ponens is well known to rule out naive truth by Curry's paradox. But there is actually no problem: in the semantics to be introduced, Weak Centering at *all* worlds guarantees only that the inference from $C \triangleright A$ and $C \triangleright (A \triangleright B)$ to $C \triangleright B$ will hold *for 2-valued sentences*; and the sentences involved in Curry-type paradoxes will not be 2-valued. (Modus Ponens, on the other hand, will be guaranteed for *all* sentences, even by Weak Centering just at normal worlds.)

[10]The canonical supposed counterexample involves a 3-candidate race whose leading candidates are a Democrat and a Republican, with an Independent far behind. Then the claim "If the Republican doesn't win, the Independent will" seems false. But "The Democrat won't win" may be true, and "If the Democrat doesn't win, then if the Republican doesn't win the Independent will win" may seem true; and these two claims lead to the false claim by Modus Ponens. A standard resolution of this, which I support, is that the complex conditional that "may seem true" isn't: what's true is only that if the Democrat doesn't win *and* the Republican doesn't win then the Independent will win, but to get from that to the complex conditional one needs the rule of Exportation $(A \wedge B) \triangleright C \models A \triangleright (B \triangleright C)$, which is invalid on the variably-strict semantics.

[11]It's also possible to add "purely modal" conditions, not involving the $\leq_w$; e.g.

S4: if $x \in W_w$ and $y \in W_x$ then $y \in W_w$.

What follows is neutral on these as well.

The first version is 2-valued:

Burgess evaluation procedure:

- $|p(c_1, ..., c_k)|_w$ is just $p_w(o_1, ...o_k)$, where $c_1, ..., c_k$ are the names for $o_1, ..., o_k$ respectively.
- $|\neg A|_w$ is $1-|A|_w$
- $|A \wedge B|_w$ is $\min\{|A|_w, |B|_w\}$
- $|\forall x A|_w$ is $\min\{|A(c/x)|_w :$ all c that name members of $U_w\}$
- $|\Box A|_w$ is $\min\{|A|_x : x \in W_w\}$
- $|A \rhd B|_w = \begin{cases} 1 & \text{iff } (\forall x \in W_w)[|A|_x = 1 \supset \\ & (\exists y \leq_w x)[|A|_y = 1 \wedge (\forall z \leq_w y)(|A|_z = 1 \supset |B|_z = 1)]] \\ 0 & \text{iff } (\exists x \in W_w)[|A|_x = 1 \wedge \\ & (\forall y \leq_w x)[|A|_y = 1 \supset (\exists z \leq_w y)(|A|_z = 1 \wedge |B|_z = 0)]] \end{cases}$

(Let a *w-neighborhood* be a non-empty subset N of W_w such that if $x \in N$ and $y \leq_w x$ then $y \in N$; and call a w-neighborhood *A-consistent* if it contains a world where $|A|$ is 1. Then the right hand side of the 1-clause for $\rhd$ says that all A-consistent w-neighborhoods have A-consistent sub-w-neighborhoods throughout which if $|A|$ is 1, so is $|B|$; and the right hand side of the 0-clause says that there is an A-consistent w-neighborhood such that every A-consistent sub-w-neighborhood of it contains a world where $|A|$ is 1 and $|B|$ is 0. If one were to make the "No Incomparabilities" assumption (for all worlds, not just normal ones) one could simplify these clauses for $\rhd$ a bit: that assumption amounts to the assumption that for each w, the w-neighborhoods are nested; and given that, the 1-clause is equivalent to the claim that if there is at least one A-consistent w-neighborhood then there is one throughout which if $|A|$ is 1, so is $|B|$.)

These stipulations give every L^+-sentence a unique value in {0,1} at each world, given any 2-valued worlds model M. Conditionals don't in general contrapose, but they shouldn't: 'If Trump runs for President he won't be elected' shouldn't imply 'If Trump is elected he won't have run'.

Validity is explained as follows:

(VAL): An inference from a set Γ of L-sentences to an L-sentence B is *Burgess-valid* if for every worlds model M and every $w \in NORM_M$, if $|A|_{M,w} = 1$ for all A in Γ then $|B|_{M,w} = 1$.

(Here what counts as a worlds model depends on which structural conditions (e.g. Weak Centering) have been imposed, so (VAL) really gives a family of notions of validity. Again, the restriction to normal worlds only makes a difference when one imposes structural requirements on the normal worlds of models that don't apply to all worlds.)[12]

[12]We can generalize to the case where B and the members of Γ can contain free variables: for any model M, if f is any function assigning $\{L^+\}_M$-names to free variables, and A is any L-formula, let A^f be the $\{L^+\}_M$-sentence that results from substitution by f. Then the generalization of (VAL) is

(VAL$_{gen}$): An inference from a set Γ of L-formulas to an L-formula B is *Burgess-valid* if for every worlds model M and every f for M and every $w \in NORM_M$, if $|A^f|_{M,w} = 1$ for all A in Γ then $|B^f|_{M,w} = 1$.

We define $\vee$ from $\wedge$ and $\neg$, and $\exists$ from $\forall$ and $\neg$, and $\Diamond$ from $\Box$ and $\neg$, in the usual ways. ($|\Diamond A|_w$ is thus $\max\{|A|_x : x \in W_w\}$.)[13] $A \lhd\rhd B$ will abbreviate $(A \rhd B) \wedge (B \rhd A)$.

But there might be a reason to treat '$\rhd$' slightly differently. Many people, myself included, find it natural to suppose that $\neg(A \rhd B)$ should be to equivalent to $A \rhd \neg B$, modulo the assumption $\Diamond A$ (that is, each should imply the other on that assumption). We don't have that on the above semantics, unless we add strong assumptions (viz.: No Ties, No Incomparabilities and the Limit Condition); that was one of Stalnaker's arguments for imposing those assumptions. If we want that equivalence without the strong assumptions, we can get it by strengthening the 0 clause for '$\rhd$' while leaving the 1 clause as is. We then need a 3-valued framework to handle sentences that receive neither value 1 nor value 0. Our worlds models are still 2-valued for the moment, i.e. atomic sentences of L^+ can only take values in $\{0,1\}$, but we allow an additional value $\frac{1}{2}$ for conditionals and sentences containing them as components. The evaluation clause for '$\rhd$' is as follows:

Modified Burgess evaluation procedure: [14]

$$|A \rhd B|_w = \begin{cases} 1 & \text{iff } (\forall x \in W_w)[|A|_x = 1 \supset (\exists y \leq_w x)[|A|_y = 1 \wedge \\ & (\forall z \leq_w y)(|A|_z = 1 \supset |B|_z = 1)]] \\ 0 & \text{iff } (\forall x \in W_w)[|A|_x = 1 \supset (\exists y \leq_w x)[|A|_y = 1 \wedge \\ & (\forall z \leq_w y)(|A|_z = 1 \supset |B|_z = 0)] \wedge (\exists x \in W_w)(|A|_x = 1) \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

(The 0-clause says that there are A-consistent w-neighborhoods, and each such has A-consistent sub-w-neighborhoods throughout which if $|A|$ is 1 then $|B|$ is 0.) I've already written the evaluation clauses for $\neg$, $\wedge$, $\forall$ and $\Box$ in a way that carries over automatically to allow for the extra value. (These clauses are called the *Strong Kleene rules.*)

The crucial thing about this alternative evaluation procedure for $\rhd$ is that if $|\Diamond A|_w$ is 1, i.e. if $(\exists x \in W_w)(|A|_x = 1)$, then $|\neg(A \rhd B)|_w$ is just $|A \rhd \neg B|_w$. Of course, a consequence will be a minimal non-classicality: excluded middle can fail for sentences containing '$\rhd$'. The cost of this isn't that high, I think: indeed, once we introduce a truth predicate, we'll need excluded middle to fail even more broadly than that.

What notion of validity goes with this modified evaluation scheme? There are several possible choices, but the one I will work with carries over the wording of (VAL) (or more generally, the (VAL$_{gen}$) of note 12) to the 3-valued case: validity involves preservation of value 1 at all normal worlds in all models (with the values now given by the modified evaluation rules).

In adding 'True' to the language we will need to adapt either the original Burgess semantics or the modified Burgess semantics to 3-valued worlds models. A *3-valued*

[13]Why take '$\Box$' as primitive, since $\Box A$ is equivalent to $\neg A \rhd A$? The answer is that the equivalence will be lost once we move to a 3-valued semantics, either because of the move to the modified Burgess evaluation procedure to be given next or to handle predicates like 'True'.

[14]There's no danger of this requiring that the same conditional get both value 0 and 1 at a world. For assume as an induction hypothesis that A and B each have a unique value at each world. (Actually we'll only need that B does.) If $A \rhd B$ gets value 0 at w, then there must be a $y \in W_w$ for which $|A|_y$ is 1 and $(\forall z \leq_w y)(|A|_z = 1 \supset |B|_z = 0)$; and if it gets 1 there must be a $y* \leq_w y$ such that $|A|_{y*} = 1 \wedge (\forall z \leq_w y*)(|A|_z = 1 \supset |B|_z = 1)$. But these require that $|B|_{y*}$ is both 0 and 1, contrary to the induction hypothesis.

worlds model is just like a 2-valued one except that in clause (iv) we replace $\{0, 1\}$ with $\{0, \frac{1}{2}, 1\}$, so that atomic sentences as well as conditionals can receive value $\frac{1}{2}$.[15] So 2-valued models are a special case of 3-valued. The most straightforward adaptation to the presence of 'True' would be to simply use the Burgess or modified Burgess rules as written. But instead of doing precisely that I will proceed in a more roundabout way, which nonetheless is modeled on these rules and agrees with them entirely for conditionals whose antecedent and consequent don't contain 'True'.

Validity will be defined as before: preservation of value 1 at all normal worlds of all models (that meet whatever structural conditions such as Weak Centering that one has imposed). However, when L contains a truth predicate we'll restrict the models used in the definition, to "arithmetically standard" models that treat the predicate 'True' in a certain way. The details are in the next section.

3. Truth and satisfaction: the strategy

Suppose that L contains a truth predicate (more specifically, a predicate of truth in L).[16] To be of interest, L will also need to have the resources to talk of the bearers of truth, i.e. sentences, and their syntactic properties. Or instead of syntactic objects, L could just contain arithmetic; we could talk of truth relative to a Gödel numbering. A language with a satisfaction predicate (from which truth can be defined, but not in general conversely) is more interesting; but to have a useful satisfaction predicate we need to be able talk of finite sequences of arbitrary objects from the universe of discourse, which requires additional mathematical resources. Moreover, dealing with satisfaction involves some notational complexity that can be confusing. So to keep things simple I'll take L to involve a truth predicate but not a satisfaction predicate. It is routine to generalize what follows from truth to satisfaction (when the extra mathematical resources are available in L).

Rather than building syntactic notions into L, I'll follow the Gödel numbering route: L will contain the predicates 'natural number', 'is zero', 'is the successor of', 'is the product of', and '$=$'. (I'll fix a Gödel numbering g of L.) I'll also be concerned only with worlds models M whose arithmetic part is standard (an ω-model) and the same from world to world. That is, I'll assume that in every model and every world in it, U_w is a superset of the set N of natural numbers, and 'natural number' is assigned N as its extension, and the other arithmetic vocabulary is interpreted in the standard way. I'll call worlds models meeting these restrictions *arithmetically standard*. It's natural to restrict to them since without some such restriction the Gödel numbering results in "non-standard syntactic expressions" that have infinitely many distinct sub-expressions. If in defining validity we restrict to arithmetically standard worlds models, the result is *ω-validity* (or *validity in ω-logic*); it is this rather than regular validity that I will be primarily concerned with.

Kripke 1975, at least the part dealing with the Kleene construction, was concerned with the possibilities for *naive* truth (and satisfaction), though in languages not containing $\rhd$. Here I will extend his results to languages containing $\rhd$.

---

[15]For present purposes I keep the earlier restriction on the assignment $=_w$ to '$=$', though with the third value it could be liberalized somewhat to allow for indeterminate identity.

[16]Not in L^+: the new names in L^+ aren't part of the language L for which we're giving a truth theory, and are dependent on a particular model of L. Any apparent loss in restricting truth to L-sentences should be met by generalizing from truth to satisfaction, as discussed later in the paragraph.

I informally defined "naive theory of truth" in my introductory remarks, but I should be more precise. Let a formula Y be *a Tr-equivalent of* a formula X if there are (possibly multiple) L-sentences A such that Y results from X by (possibly multiple) substitutions of $True(\langle A\rangle)$ for A and/or vice versa. A *naive theory of truth* is one where whenever Y is a Tr-equivalent of X, Y follows from X and vice versa (i.e. the inferences from X to Y and Y to X are valid). The semantic paradoxes show that naivety is unattainable in classical logic, but Kripke (in his Kleene-based construction) showed it attainable in non-classical logic, by the use of 3-valued models. (Again, his language didn't contain $\triangleright$.)

Naivety is not the sole requirement we should impose on a theory of truth: we also want it to obey reasonable compositional laws, and to allow the truth predicate to appear in an induction rule. More on these shortly.

Our theory of truth should of course also be consistent, at least Post-consistent: that is, it shouldn't imply everything. I don't *in principle* require negation-consistency, i.e. the restriction to theories that for no A imply both A and $\neg A$. However, as is implicit in my earlier definition of validity, the theories I'll be developing satisfy disjunctive syllogism ($A \vee B, \neg A \vDash B$), and for those theories Post-consistency requires negation-consistency. (While there are familiar "paraconsistent" logics that avoid paradoxes without restricting excluded middle, by restricting disjunctive syllogism instead, they don't seem to me a promising framework for my ultimate goal of restricted quantification: the comments in Section 7 below on Beall *et al* 2006 and Beall 2009 may be enough to give some sense of this.)

Actually we want our naive truth theory to be more than (Post- or negation-) consistent: a consistent theory might, after all, imply the defeat of the Paris Commune, and no logic of truth should do that. What we want is for our theory of truth to be "consistent with any arithmetically standard worlds model" of the 'True'-free fragment of L, which I'll call L_0. More fully,

GOAL: We want to generate from each 2-valued arithmetically standard worlds model M_0 for L_0 a corresponding 3-valued worlds model M for L that (a) validates naive truth and (b) is exactly like M_0 except that it assigns a 3-valued extension to 'True'. It follows from (b) that the sentences of L_0^+ get the same value at w in M as in M_0, for each world w; and also that M is arithmetically standard, given that M_0 is.

I'll take the allowable worlds models M of L to be just the ones generated from worlds models M_0 of L_0 in this way; that is, validity, consistency etc. *in the logic of truth* are defined by quantification over the arithmetically standard worlds models M_0 of the 'True'-free fragment of the language, and extending the valuation to sentences with 'True' by a procedure to be given.[17] (It isn't immediately obvious what this procedure should be when it comes to sentences containing both '$\triangleright$' and 'True': e.g. to take a very simple Curry-like case, it isn't immediately obvious how to evaluate a sentence $K_\triangleright$ constructed by the usual Gödel-Tarski techniques so as to be equivalent to $True(\langle K_\triangleright\rangle) \triangleright \neg True(\langle K_\triangleright\rangle)$. Indeed I will consider several alternative procedures for constructing the extension.)

Note that if we can establish (GOAL), we get a kind of conservativeness result: letting *-consistency be consistency in ω-logic, we have that any classically

[17] A slightly more general procedure will be mentioned in note 30.

*-consistent set of sentences of L_0 is *-consistent in a naive truth theory.[18] The naive truth theory in question includes not merely the inferences from any sentence to its Tr-equivalents, it can include any other law validated in the construction of M from M_0. What laws these are will of course depend on the details of the construction of M from M_0, which is yet to be given.

But whatever the details, it is clear in advance that if (GOAL) is achieved then our construction will not only be one on which truth is naive, but one where mathematical induction in the form $A(0) \wedge (\forall n \in N)(A(n) \supset A(n+1)) \models (\forall n \in N)A(n)$ is legitimate even for formulas containing 'True'.[19] The reason is that in any arithmetically standard worlds model, when the premises of this induction rule hold at a world the conclusion must too, and the construction guarantees that the new worlds model is arithmetically standard.

It is almost as immediate that the construction will validate the desirable composition principles, e.g.

COMPOS-GENERAL: $\forall x \forall y \forall z$(If x and y are sentences and z is the result of applying '$\triangleright$' to x and y in that order, then $\Box[True(z)$ if and only if $(True(x) \triangleright True(y)]$).

For as long as the logic validates each instance of "$\Box$ [A if and only if A]", then the naivety of truth guarantees the validity of each instance of

COMPOS-SCHEMA: $\Box[True(\langle A \triangleright B \rangle)$ if and only if $(True(\langle A \rangle) \triangleright True(\langle B \rangle))]$;

and since the constructed model is arithmetically standard, the generalization is guaranteed to hold in the model when the instances do. [This holds on *any* reading of 'if and only if', as long as "$\Box$ [A if and only if A]" is validated. At the moment, the only available reading is '$\triangleleft\triangleright$', but I will later add other biconditionals, and the point applies equally to them.]

4. Truth and satisfaction: the details

I now outline a generalization of Kripke's construction. The initial generalization, which takes '$\triangleright$' as a black-box, is completely routine, hardly a generalization at all; but a non-Kripkean ingredient is then required, to give a substantial account of '$\triangleright$'.

Let's get the pure Kripke part of the construction out of the way first. It's clear from what has already been said that each of the worlds w in the model for L will be evaluated in part on the basis of U_w and the w-extensions of L_0-predicates. The additional ingredients needed to evaluate L^+-sentences at each w are:

- a 3-valued extension T_w for 'True': it assigns values in $\{0, \frac{1}{2}, 1\}$ to objects in U. (We'll want it to assign non-zero values only to those objects that are Gödel numbers of L-sentences under the chosen Gödel numbering.)
- a function j_w that assigns to each L^+-sentence of form '$A \triangleright B$' a value in $\{0, \frac{1}{2}, 1\}$.

Let T and j be the functions that assign to each $w \in W$ a T_w and j_w. Relative to any such T and j, the Kleene rules tell us how to evaluate every L^+-sentence at w:

[18] Calling this a conservativeness result could be misleading: there is no *deductive* conservativeness, it is a kind of semantic conservativeness in ω-logic. Its purpose, as I've said, is to ensure that the set of principles to be declared valid in the naive truth theory is not merely consistent, but consistent with any set of assumptions in the 'True'-free language that are compatible with the conditional logic and standard models of arithmetic.

[19] Analogous forms with other modus-ponens obeying conditionals in place of the '$\supset$' are guaranteed too.

- For p other than 'True', $|p(c_1, ..., c_k)|_{w,j,T}$ is just $p_w(o_1, ...o_k)$;
- $|True(c)|_{w,j,T}$ is $T_w(o)$, where o is the object denoted by the L^+-name c;
- $|\neg A|_{w,j,T}$ is $1 - |A|_{w,j,T}$
- $|A \wedge B|_{w,j,T}$ is min $\{|A|_{w,j,T}\ ,\ |B|_{w,j,T}\ \}$
- $|\forall x A|_{w,j,T}$ is min$\{|A(c/x)|_{w,j,T}$: all c that name members of $U_w\}$
- $|\Box A|_{w,j,T}$ is min $\{|A|_{w,j,T} : x \in W_w\}$
- $|A \rhd B|_{w,j,T} = j_w(A \rhd B)$.

The important thing about this is a monotonicity principle. Let $T \leq_K T^*$ mean that for every w and every L-sentence S, if $T_w(S) = 1$ then $T^*_w(S) = 1$ and if $T_w(S) = 0$ then $T^*_w(S) = 0$. Then

(MONOT): For any M and j: if $T \leq_K T^*$ then for any $w \in W$ and any L^+-sentence A, if $|A|_{w,j,T} = 1$ then $|A|_{w,j,T^*} = 1$ and if $|A|_{w,j,T} = 0$ then $|A|_{w,j,T^*} = 0$.

This is easily proved by an induction on the complexity of A. (The result is familiar from Kripke 1975, except that I've added a trivial $\rhd$ clause and a world-argument for T.)

This is the background for

Proposition. ***[Kripke's observation.]*** *For any M and j, there are T ("Kripke fixed points" relative to M and j) for which, for each $w \in W$:*

For every L-sentence A, $|A|_{w,j,T} = T_w(g(A))$ [and hence $|A|_{w,j,T} = |True(c)|_{w,j,T}$, where c denotes $g(A)$]; and

$T_w(o)$ is 0 if o is not $g(A)$ for some L-sentence A.

In particular, for any M and j there is a minimal fixed point T_{min}, i.e. a fixed point (relative to M and j) such that for every other fixed point T (relative to M and j), $T_{min} \leq_K T$.

Kripke's observation is easily proved by transfinite induction.[20]

It easily follows that as long as j is *transparent*, in the sense that it assigns Tr-equivalent formulas the same value, then the naivety condition is met: whenever A and B are Tr-equivalent, $|A|_{w,j,T_{min}} = |B|_{w,j,T_{min}}$. (And similarly for fixed points T other than T_{min}.)

The definition of T_{min} depended on the choice of M and j, but given those, T_{min} is uniquely determined; so we can abbreviate $|A|_{w,j,T_{min}}$ as $|A|_{w,j}$. To repeat, this valuation yields naive truth as long as j is transparent.

The harder task is to construct an appropriate transparent j-function for evaluating conditionals at worlds. What we want is a transparent j that leads to a logic that reduces to the Burgess or modified-Burgess logic when applied to 'True'-free sentences *and which weakens the laws as little as possible when sentences with 'True' are allowed as instances.* There are at least two approaches to constructing

[20]Holding M and j fixed, we define T_0 to be the function assigning the value $\frac{1}{2}$ to every Gödel number of an L-sentence, and 0 to everything else; $T_{\sigma+1}$ the function assigning every world w and L-sentence A the value $|A|_{w,j,T_\sigma}$; and T_λ (for limit λ) the function assigning every world w and L-sentence A the value

$$\begin{cases} 1 & \text{if for some } \sigma<\lambda \text{ and every } \tau \text{ such that } \sigma \leq \tau<\lambda,\ |A|_{w,j,T_\tau} = 1; \\ 0 & \text{if for some } \sigma<\lambda \text{ and every } \tau \text{ such that } \sigma \leq \tau<\lambda,\ |A|_{w,j,T_\tau} = 0; \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

We can then easily prove by induction that if $\sigma < \tau$, $T_\sigma \leq_K T_\tau$. Cardinality considerations then show that there are ordinals σ (of the cardinality of U_M) after which the assigned T never changes. Taking T_{min} to be T_σ for such a σ, we get the desired result.

such a j function: a revision construction, with similarities to those in Field 2008; or a fixed point construction, with similarities to those in Field 2014.

The revision construction is simpler, so I'll focus on it, but will also make a few remarks about the (perhaps more aesthetically pleasing) fixed point construction.

4.1. **The revision construction.** Fix a worlds model M_0 for L_0. Suppose we have given a provisional valuation j_ν, which assigns values $|B \triangleright C|_{w,j_\nu}$ to any L^+-sentences B and C. As we've seen, this indirectly gives a value $|A|_{w,j_\nu}$ to every L^+-sentence A at every world, via the Kripke minimal fixed point construction; let's just write this as $|A|_{w,\nu}$. We want to use this valuation j_ν to construct a revised one $j_{\nu+1}$, perhaps a better one, which is transparent if the original one is; the structure of worlds is used in the revision.

There are two possibilities for $j_{\nu+1}$, one based on the original Burgess valuation rules and the other based on the variant. For the original it is:

$$j_{w,\nu+1}(A \triangleright B) \text{ is } \begin{cases} 1 & \text{iff } (\forall x \in W_w)[|A|_{x,\nu} = 1 \supset (\exists y \leq_w x)[|A|_{y,\nu} = 1 \wedge \\ & (\forall z \leq_w y)(|A|_{z,\nu} = 1 \supset |B|_{z,\nu} = 1)]] \\ 0 & \text{iff } (\exists x \in W_w)[|A|_{x,\nu} = 1 \wedge (\forall y \leq_w x)[|A|_{y,\nu} = 1 \supset \\ & (\exists z \leq_w y)(|A|_{z,\nu} = 1 \wedge |B|_{z,\nu} = 0)]] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

For the variant, it's the same except for a modified 0 clause:

$$0 \text{ iff } (\forall x \in W_w)[|A|_{x,\nu} = 1 \supset (\exists y \leq_w x)[|A|_{y,\nu} = 1 \wedge (\forall z \leq_w y)(|A|_{z,\nu} = 1 \supset |B|_{z,\nu} = 0)]] \wedge (\exists x \in W_w)(|A|_{x,\nu} = 1).$$

Choose whichever you like: the construction that follows works with either choice.

To get the revision process started, we need a starting valuation j_0, and we want it to be transparent since this will guarantee that later j_ν are as well. For simplicity I'll take a trivial j_0, which assigns value $\frac{1}{2}$ to each conditional at each world. It makes little difference, because the effect of the starting values gets *almost* completely wiped out as the construction proceeds. (It gets *completely* wiped out for sentences not containing 'True': whatever the starting values, any such sentence gets the value that it gets in the 2-valued worlds model for the corresponding version of Burgess semantics by stage n, where n is the maximum depth to which '$\triangleright$' is embedded in the scope of other '$\triangleright$'s in A; and it keeps that value at all subsequent stages. So from stage ω on, all 'True'-free sentences get "the value they should", whatever the starting valuation.)

Finally, we need a policy on limit stages. Here the choice is important, and we choose continuity with respect to 1 and 0. That is, if λ is a limit ordinal then for any world w and any conditional $A \triangleright B$, j_λ assigns the conditional 1 at a world if and only if for some $\mu < \lambda$, for every ordinal ν in the open interval (μ, λ) assigns the conditional value 1 at that world; and similarly for 0. (So "irregularity arbitrarily close to λ" at a world as well as "constant $\frac{1}{2}$ sufficiently close to λ" at that world lead to value $\frac{1}{2}$ at λ at that world.)

We can summarize these choices in a single definition. For the semantics based on the modified Burgess, which I prefer, it's

$$j_{w,\kappa}(A \triangleright B) \text{ is } \begin{cases} 1 & \text{if } (\exists\mu < \kappa)(\forall\nu \in [\mu,\kappa))(\forall x \in W_w)[|A|_{x,\nu} = 1 \supset (\exists y \leq_w x) \\ & [|A|_{y,\nu} = 1 \wedge (\forall z \leq_w y)(|A|_{z,\nu} = 1 \supset |B|_{z,\nu} = 1)]] \\ 0 & \text{if } (\exists\mu < \kappa)(\forall\nu \in [\mu,\kappa))[(\forall x \in W_w)[|A|_{x,\nu} = 1 \supset (\exists y \leq_w x) \\ & [|A|_{y,\nu} = 1 \wedge (\forall z \leq_w y)(|A|_{z,\nu} = 1 \supset |B|_{z,\nu} = 0)]] \wedge (\exists x \in W_w)(|A|_{x,\nu} = 1)] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

($[\mu,\kappa)$ is the half-open interval of ν such that $\mu \leq \nu < \kappa$.) For the semantics based on the original Burgess, modify the 0 clause in the obvious way.

It's evident that on either variant, each j_κ is transparent if all preceding j_ν are transparent; so by transfinite induction, all are transparent.

At each world, all 'True'-free sentences get the desired value (i.e. the one given in the 2-valued model from which we started) by stage ω, and keep it at later stages. But there is much greater irregularity for sentences containing 'True', due to the interaction between 'True' and '$\triangleright$'.[21] In particular there is no fixed point. How then are we to select a privileged j?

The sequence of j_ν is a revision sequence in the sense of Gupta and Belnap 1993. (The revision sequence depends on the model M_0, as well as on the choice of Burgess or modified-Burgess.) One well-known feature of revision sequences is that there are evaluations j that appear arbitrarily late in the revision process; indeed, there are ordinals κ such that for any $\mu \geq \kappa$ and any ζ, there is a $\nu \geq \zeta$ such that $j_\nu = j_\mu$.[22] Call any infinite such κ *final* (relative to model M_0),[23] and let FIN (or FIN_{M_0}) be the class of final ordinals.

But not all final ordinals assign the same j (if they did, it would be a fixed point). Which to pick? Obviously we want one that will yield as nice laws for $\triangleright$ as possible. Gupta and Belnap 1993 have a general theorem, their Reflection Theorem, that we can bring to bear. Applied to this case, that theorem says:

Proposition. ***[Gupta-Belnap]*** *There are limit ordinals Ω ("reflection ordinals for the sequence j_κ")*[24] *such that*

(i) Ω is final

(ii) For any L^+-formulas A and B, and any world w and any $d \in \{0, \frac{1}{2}, 1\}$,

$(\exists\mu < \Omega)(\forall\nu \in [\mu,\Omega))((j_{w,\nu}(A \triangleright B) = d)$ *if and only if* $(\forall\nu \in FIN)(j_{w,\nu}(A \triangleright B) = d)$.

Moreover, in the above semantics these reflection ordinals have an especially useful property:

Proposition. ***[Fundamental Theorem for L (revision-theoretic version).]*** *For any reflection ordinal Ω, any $w \in W$, and any L^+-sentence A,*

(a) $|A|_{w,\Omega} = 1$ *if and only if* $(\forall\nu \in FIN)(|A|_{w,\nu} = 1)$

and *(b)* $|A|_{w,\Omega} = 0$ *if and only if* $(\forall\nu \in FIN)(|A|_{w,\nu} = 0)$.

---

[21] The sentence itself needn't even contain '$\triangleright$' for the irregularity to occur, because the use of 'True' typically makes other sentences relevant to the evaluation.

[22] Since the revision sequence here is *Markovian* in the sense that for any ordinals μ, κ and ν, if $j_\mu = j_\kappa$ then $j_{\mu+\nu} = j_{\kappa+\nu}$, we can simplify to: for any ζ, there is a$\nu \geq \zeta$ such that $j_\nu = j_\kappa$. If this holds for κ in a Markovian sequence, it is bound to hold for any $\mu > \kappa$.

[23] It isn't really necessary to demand infinitude explicitly, it's entailed by the rest, as the reader can easily prove using 'True'-free sentences where '$\triangleright$' is embedded to depth n for arbitrarily large n.

[24] Which ordinals are reflection ordinals will depend on the starting model M_0.

Since there is only one possible value other than 0 and 1, these two clauses imply that each reflection ordinal Ω is associated with the same j_Ω. This j_Ω is the valuation for $\triangleright$-conditionals that I'll be employing, e.g. in determining validity.

The Fundamental Theorem as stated here is similar to that given in Field 2008, but the conditional there was different. The proof given there included a proof of [Gupta-Belnap], since I was unaware of their theorem at the time. (Belated apologies to them for not being able to give credit.) A proof of the Fundamental Theorem for the language of this paper, now relying on [Gupta-Belnap] to save work, is given in Appendix A.

Note that when A is a conditional $B \triangleright C$, the 1-clause of the Fundamental Theorem together with the evaluation rules for $\triangleright$ yield that for any reflection ordinal Ω and $w \in W$,

1-clause: $|B \triangleright C|_{w,\Omega} = 1$ if and only if $(\forall \nu \in FIN)(\forall x \in W_w)[|B|_{x,\nu} = 1 \supset (\exists y \leq_w x)[|B|_{y,\nu} = 1 \wedge (\forall z \leq_w y)(|B|_{z,\nu} = 1 \supset |C|_{z,\nu} = 1)]]$.

Since $\Omega \in FIN$, this yields a necessary but not sufficient condition for $|B \triangleright C|_{w,\Omega} = 1$ that involves no ordinals other than Ω:

1-clause Corollary: If $|B \triangleright C|_{w,\Omega} = 1$ then $(\forall x \in W_w)[|B|_{x,\Omega} = 1 \supset (\exists y \leq_w x)[|B|_{y,\Omega} = 1 \wedge (\forall z \leq_w y)(|B|_{z,\Omega} = 1 \supset |C||_{z,\Omega} = 1)]]$.

That is, since we've chosen to use j_Ω for our final valuation: the 1-clause we've adopted is strictly stronger than the 1-clause of the Burgess and modified-Burgess semantics. But since all final ordinals are infinite, all 'True'-free sentences receive the same value at all final ordinals; this means that for such B and C the 'if...then' in the corollary becomes an 'if and only if'. In other words, we're guaranteed that *the Burgess/modified-Burgess 1-clause is retained for 'True'-free sentences.*

Moreover, as long as we have Weak Centering at w, the 1-clause corollary yields the following for all B and C (not just the 'True'-free ones):

Modus Ponens for $\triangleright$: If $|B \triangleright C|_{w,\Omega} = 1$ and $|B|_{w,\Omega} = 1$ then $|C|_{w,\Omega} = 1$.

(The label 'Modus Ponens' is really appropriate only if we have Weak Centering at all normal w.)

Something similar holds for the 0-clause, though the details depend on which version of the 0 clause one uses. In both cases, we get strictly stronger conditions than would be given by direct application of the Burgess or modified Burgess rules: e.g. for the semantics based on modified Burgess we get

If $|B \triangleright C|_{w,\Omega} = 0$ then $(\forall x \in W_w)[|B|_{x,\Omega} = 1 \supset (\exists y \leq_w x)[|B|_{y,\Omega} = 1 \wedge (\forall z \leq_w y)(|B|_{z,\Omega} = 1 \supset |C|_{z,\Omega} = 0)]] \wedge (\exists x \in W_w)(|B|_{x,\Omega} = 1)$.

But again, when confined to 'True'-free sentences the 'if' becomes an 'if and only if': *the Burgess or modified Burgess 0 clause is also retained for 'True'-free sentences.*

(When w is weakly centered, the above yields

0 Law for $\triangleright$: If $|B \triangleright C|_{w,\Omega} = 0$ and $|B|_{w,\Omega} = 1$ then $|C|_{w,\Omega} = 0$ (and indeed, $|C|_{x,\Omega} = 0$ whenever $x \sim_w w$),

which also strikes me as desirable but will play no role in what follows. Had we based the semantics on the original Burgess, we'd have needed that w be strongly centered to get this result.)

4.2. **Where are we?** For each starting arithmetically standard worlds model M_0 for the 'True'-free fragment L_0 of L (with $\triangleright$ evaluated either by the standard Burgess or variant Burgess rules), we have chosen a transparent j_Ω to evaluate all L^+-conditionals at each world (including those containing embedded conditionals and/or 'True'), and a T to evaluate truth-claims at each world. The worlds, and their division into normal and non-normal, are the same in the new model as in the old. (In particular, if the old contains no non-normal worlds, the new one won't either.) The assignment of accessibility sets W_w and pre-orders $\leq_w$ is also the same in the new model as in the old; so are the assignments of extensions to predicates at each world. And at each world, j_Ω assigns the same values to 'True'-free conditionals (and hence 'True'-free sentences more generally) as the original model on M_0 did. Finally, by the transparency of j and the features of the Kripke construction, the truth predicate is naive; and since the model is arithmetically standard, there can be no worry about using formulas with 'True' in the induction rule or validating generalities (e.g. composition rules) whose instances are valid.[25]

The following are laws of this construction: by which I mean, schemas all of whose instances are valid (whatever structural conditions, such as Weak Centering at normal worlds, we decide on):

- $A \triangleright A$
- $[A \triangleright (B \wedge C)] \triangleleft\triangleright [(A \triangleright B) \wedge (A \triangleright C)]$
- $[(A \triangleright C) \wedge (B \triangleright C)] \triangleright [(A \vee B) \triangleright C]$
- $[A \triangleright (B \wedge C)] \triangleright [(A \wedge B) \triangleright C]$.

These are laws both when the evaluation rule for $\triangleright$ is based on the original Burgess rule and when it is based on the modified rule: the 0 clause makes no difference. Indeed on both constructions they are all *strong laws*, by which I mean that their instances have value 1 at all worlds of every model, not just all normal worlds. That's important because it means that the result of prefixing any string of $\Box$s and $\Diamond$s to one of these is also a law. Related, it guarantees other "regular behavior", such as that we can strengthen antecedents in the laws. That is, even though we don't want and don't get that $Y \triangleright Z$ entails $X \wedge Y \triangleright Z$ for variably strict conditionals, still if $Y \triangleright Z$ is a strong law then so is $X \wedge Y \triangleright Z$ (even if X is true only at non-normal worlds). Similarly, if $X \triangleright Y$ and $Y \triangleright Z$ are strong laws then so is $X \triangleright Z$.[26] Proving that the bulleted schemas are strong laws is straightforward.[27] Note that since $\Box(A \triangleleft\triangleright A)$

---

[25] A feature of the model as described is that it is not value-functional: the value of $A \triangleright B$ at a world isn't determined wholly by the values of A and B at it and other worlds. The reason is that all these values are values at a reflection Ω, and these depend on values at all non-reflection ordinals in FIN. But it isn't hard to use what's been done here to construct an enriched value space (along the lines of Field 2008, Section 17.1) in which we do have value-functionality: the value space for that will have infinitely many values, not linearly ordered. (The space is a set of functions from an initial segment of the ordinals to $\{0, \frac{1}{2}, 1\}$, where the length of the initial segment is the distance between successive reflection ordinals.) But for purposes of this paper there's no need for value-functionality.

[26] The proof that "antecedent strengthening" and transitivity are legitimate for strong laws uses the Fundamental Theorem as applied to $\triangleright$-sentences. Let W^* be the set of worlds that are n-accessible from worlds for some n. (On reasonable assumptions this will just be W, but the proof doesn't need this.) For antecedent strengthening, suppose that $Y \triangleright Z$ has value 1 at all worlds at reflection ordinals. Then it has value 1 at all worlds at all final ordinals, which means that at all final ordinals and all worlds in W^*, if Y has value 1 then so does Z; and that includes all worlds where X has value 1. From this it's evident that $X \wedge Y \triangleright Z$ has value 1 at all worlds in W (even those not in W^*, since only those in W^* are accessible to them) at all final ordinals, and in particular at reflection ordinals. The argument for transitivity is similar.

[27] The key observation for all of them is that for $|X \triangleright Y|_{w,\Omega}$ to be 1, it suffices that for all worlds w^* and all final ordinals ν, if $|X|_{w^*,\nu} = 1$ then $|Y|_{w^*,\nu} = 1$. Given that, it's simply a matter of relativizing the proof that one would give for the Burgess-based semantics in the ground

is valid, then by naivety so are $\Box(True(\langle A\rangle) \triangleleft\triangleright A)$ and $\Box(\neg True(\langle A\rangle) \triangleleft\triangleright \neg A)$, and hence $\Box(True(\langle \neg A\rangle) \triangleleft\triangleright \neg True(\langle A\rangle))$. (And by the remarks at the end of Section 3, this means that we have a general composition principle for negation: for any sentence x, the negation of x is true if and only if x is not true.)

The fact that the above laws all hold in the construction with naive truth is interesting, because these are exactly the axiom schemas that Burgess uses in the quantifier-free case for the 'True'-free fragment of the language. He gives a completeness proof there, for a system with these axioms, a necessitation rule, and the rule that for any string P of $\Box$s and $\Diamond$s, if $\vDash P\Box(A \equiv B)$ then $\vDash P[(A\triangleright C) \equiv (B\triangleright C)]$. The last rule is inappropriately weak in the 3-valued framework: we want a rule that has bite even when A and B aren't bivalent. (An adequate replacement requires the additional conditional '$\rightarrow$' soon to be introduced).[28] More generally, because the 3-valued background is weaker, the Burgess axiomatization doesn't give a complete proof-procedure in the 3-valued context.[29] Still, I think that the fact that his axioms carry over unchanged is some indication that adding a naive truth predicate hasn't seriously compromised the laws of '$\triangleright$' (and once we add the '$\rightarrow$' things will look even better).

In addition, we've seen that as long as we restrict the ground models to those with Weak Centering at normal worlds (as is required for Modus Ponens in the ground language), then Modus Ponens for $\triangleright$ also holds in the expanded logic with 'True'. (Some of the laws obtained in the 2-valued logic by adding restrictions on the $\leq_w$ can only be carried over straightforwardly to the full logic with 'True' when stated using the aforementioned conditional '$\rightarrow$' that generalizes the material conditional. We'll turn to that conditional in Section 5.)

That's the revision construction.

4.3. **The fixed point construction.** As I've mentioned, one can also give a fixed point construction that yields a rather similar outcome. Again consider valuation

level semantics to a given ν. For instance, for the right to left direction of the second listed law: Suppose that $|A \triangleright B|_{w^*,\nu} = |A \triangleright C|_{w^*,\nu} = 1$. Then for every x in W_{w^*} such that $|A|_{w^*,\nu} = 1$, there is a $y_1 \leq_{w^*} x$ such that

$$\text{(a) } |A|_{y_1,\nu} = 1 \wedge (\forall z \leq_{w^*} y_1)[|A|_{z,\nu} = 1 \supset |B|_{z,\nu} = 1],$$

and for every y_1 in W_{w^*} such that $|A|_{y_1} = 1$, there is a $y_2 \leq_{w^*} y_1$ such that

$$\text{(b) } |A|_{y_2,\nu} = 1 \wedge (\forall z \leq_{w^*} y_2)[|A|_{z,\nu} = 1 \supset |C|_{z,\nu} = 1].$$

Since $\leq_{w^*}$ is a pre-order on W_{w^*}, (a) entails its analog (a*) where y_2 replaces y_1; and that with (b) yields

$$|A|_{y_2,\nu} = 1 \wedge (\forall z \leq_{w^*} y_2)[|A|_{z,\nu} = 1 \supset |B \wedge C|_{z,\nu} = 1],$$

which entails $|A \triangleright B \wedge C||_{w^*,\nu} = 1$.

(This proof and the proofs of the other laws just given doesn't depend on the use of a reflection ordinal for our evaluation: that should be no surprise, since the Fundamental Theorem shows that a single sentence can only have value 1 at reflection ordinals if it has value 1 at all final ordinals. Where the fact that validity requires preservation of value 1 only at reflection ordinals is important is for inferences from premises: e.g. Modus Ponens (assuming Weak Centering at normal worlds) and $A \wedge B \vDash A \triangleright B$ (assuming strong).)

[28]The best replacement is:

For any string P of $\Box$s and $\Diamond$s, if $\vDash P\Box(A \leftrightarrow B)$ then $\vDash P[(A \triangleright C) \Leftrightarrow (B \triangleright C)]$;

here '$\Rightarrow$' is defined from '$\rightarrow$' and strengthens it in a way to be discussed in Section 5, and '$\leftrightarrow$' and '$\Leftrightarrow$' are defined from '$\rightarrow$' and '$\Rightarrow$' in the obvious ways. (The displayed law with mixed biconditionals entails the versions with two '$\leftrightarrow$' and with two '$\Leftrightarrow$'.)

[29]Indeed, the fact that we've restricted to arithmetically standard models immediately rules out the possibility of a complete proof procedure.

functions j that assign values in $\{0, \frac{1}{2}, 1\}$ to each pair of a world and a $\triangleright$-conditional; again we're only interested in valuation functions that are transparent. The idea is to show that there is a set **J** of transparent valuations, with a distinguished member j^*, where we have

Proposition. *[Fundamental Theorem for* L *(fixed point version).]* *For any* $w \in W$*, and any* L^+*-sentence* A*,*

(a) $|A|_{w,j^*} = 1$ *if and only if* $(\forall h \in J)(|A|_{w,h} = 1)$

(b) $|A|_{w,j^*} = 0$ *if and only if* $(\forall h \in J)(|A|_{w,h} = 0)$.

So j^* plays more or less the role that the j_Ω for reflection Ω play in the revision approach, and **J** plays more or less the role of the set of those j that occur arbitrarily late in the revision process (i.e. at ordinals in FIN). Here too, the various valuations j get a semantics whereby for any L^+-sentences A and B and any world w, $j(w, A \triangleright B)$ is determined in a natural way from the values that valuations related to j give to B in worlds near w where A has value 1; and the semantics gives the values in the original model to L^+-sentences not containing 'True'. To get the proper intersubstitutivity of logical equivalents, one needs to set up the semantics in a slightly non-obvious way. I sketch the construction in Appendix B; it is a generalization to variably strict conditionals of the one in Field 2014, and that paper will enable the reader to easily fill out the sketch in the Appendix.

(The basic idea of using a fixed point on a set of valuations was suggested in Yablo 2003; but Yablo's procedure didn't cut down the set of valuations quantified over in the semantics of each world nearly far enough—indeed, highly irregular valuations were included—and this led to extreme failure of intersubstitutivity of logical equivalents in embedded conditionals. Introducing chains in the manner of Appendix B seems to be the simplest acceptable way of accommodating Yablo's basic insight.)[30]

The remarks in Section 4.2 about the revision construction carry over to the fixed point construction virtually unchanged. In particular, the laws listed there are valid here too (again, with Modus Ponens as long as the original model has Weak Centering at normal worlds).

5. "Material-like" conditionals

Many uses of 'if ... then' in English are captured reasonably well by a variably strict conditional like '$\triangleright$', but some uses are more in line with a material conditional: in particular, the conditional used to restrict universal quantification is. "All A are B" can't be rendered as $\forall x(Ax \triangleright Bx)$: that's too strong when '$\triangleright$' is an ordinary indicative (or subjunctive) conditional. For instance, "Everyone who will be elected President in 2016 is female" might be true but "For everyone x, if x is elected President in 2016 then x is female" presumably isn't: on the ordinary indicative reading, Jeb Bush and many others are counterexamples even if unelected. In a 2-valued context, we can represent "All A are B" as $\forall x(Ax \supset Bx)$, where this is short for $\forall x(\neg Ax \vee Bx)$. But in a 3-valued context with restrictions on excluded

[30]Yablo's paper also suggests the use of multiple Kripke fixed points for 'True' instead of the minimal ones; that idea can be employed with any of the constructions for '$\triangleright$' in this section, both revision-theoretic and fixed point, and has what are arguably some advantages. For further discussion (in a revision-theoretic context with a different conditional), see Field 2008, Section 17.5. Again, it doesn't matter to the issues of this paper whether one makes these modifications.

middle, we can't use a $\supset$ defined in terms of $\neg$ and $\vee$ (at least if we want such schemas as "All A are A" and "All A are either A or B" to be logical laws); we need a new conditional '$\rightarrow$' or '$\Rightarrow$', that reduces to $\supset$ for 2-valued sentences just as our '$\triangleright$' reduces to the "classical" variably strict conditional.[31] I find it plausible that this quantifier-restricting conditional is contraposable, but I needn't insist on this: I will simply take '$\Rightarrow$' to be a contraposable conditional and '$\rightarrow$' to be a non-contraposable one, and we can leave open for now which of the two is to be used to define restricted quantification. There is no need for separate theories of '$\rightarrow$' and '$\Rightarrow$': we can take the basic conditional to be the non-contraposable '$\rightarrow$', and define $A \Rightarrow B$ as $(A \rightarrow B) \wedge (\neg B \rightarrow \neg A)$, which ensures that '$\Rightarrow$' is contraposable. The basic '$\rightarrow$' and the derived '$\Rightarrow$' have uses other than for restricting quantification: as observed in note 28, they are also needed for some of the the laws of '$\triangleright$' (and for these purposes, '$\rightarrow$' as well as '$\Rightarrow$' is required). But though I'll take '$\rightarrow$' as basic, '$\Rightarrow$' will be the primary focus, because at least in my own view, it is the contraposable one that is ordinarily used to restrict universal quantification.

There are several options in the literature for such a conditional '$\rightarrow$' (or a corresponding contraposable '$\Rightarrow$'). Some of these are broadly like the revision-theoretic and fixed point options for $\triangleright$ given in Section 4; but a key difference is that the valuations at a single world look only at other values at that same world.

For the moment let's ignore the interaction between '$\rightarrow$' and '$\triangleright$', and focus on a language L^* just like L except that it has '$\rightarrow$' instead of '$\triangleright$'. A language with both '$\rightarrow$' and '$\triangleright$' is far more interesting, and will be treated in Section 6. That is what we'll need for a proper logic for restricted quantification in naive truth theory, a matter I'll turn to in Section 7. But for the moment, I look at L^*, which has '$\rightarrow$' only.

L^*, like L, contains 'True'; if it didn't, and could be given a 2-valued semantics, we could just define $\rightarrow$ from $\neg$ and $\vee$ in the usual way. As before, the semantics for 'True' will be given by Kripkean constructions in which valuations v (analogous to the previous j) for '$\rightarrow$' at each world are held fixed; the real work then consists in the specification of an appropriate valuation for '$\rightarrow$' at each world.

A revision-theoretic construction of such a valuation for '$\Rightarrow$' was given in Field 2008; instead of what I called the "Official Conditional", given in Ch. 16, I now prefer the "first variation" given in Section 17.5, which modifies the 0 clause.[32] And I want to adapt it to the non-contraposable '$\rightarrow$'. Since L^* contains '$\Box$', we need to add a worlds parameter; but the semantics for '$\rightarrow$' is given world-by-world, unlike for '$\triangleright$', and is thus considerably simpler. It goes like this:

---

[31]I should note that the notation used in this paper is almost the reverse of the notation in Field 2014. There, the material-like conditional used to restrict quantification (which was assumed contraposable) was symbolized as $\blacktriangleright$, and $\triangleright$ was its non-contraposable generalization; whereas $\rightarrow$ was used to symbolize a conditional with very much the flavor of the $\triangleright$ used here, though it wasn't based on a Stalnaker-Lewis-Pollock-Burgess multiple worlds semantics. Sorry for any confusion, but I think the new notation distinctly better.

An alternative to introducing a new conditional and defining universal restricted quantification in terms of it is to take a binary restricted quantifier $(\forall x \ni Ax)Bx$ as primitive. One can define '$\Rightarrow$' (though not '$\rightarrow$') from it, as well as the other way around.

[32]This switch yields a cleaner relation between $|A \Rightarrow B \wedge C|$ on the one hand and $|A \Rightarrow B|$ and $|A \Rightarrow C|$ on the other: see the end of this section. That in turn is important for restricted quantifier law 4a* in Section 7.

$$|A \to B|_{w,\alpha} = \begin{cases} 1 & \text{if } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha))[|A|_{w,\gamma} = 1 \supset |B|_{w,\gamma} = 1] \\ 0 & \text{if } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha))[|A|_{w,\gamma} = 1 \wedge |B|_{w,\gamma} = 0] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

If we then define '$\Rightarrow$' from '$\to$' as above, we get something similar but with a strengthened 1-clause:

$$|A \Rightarrow B|_{w,\alpha} = \begin{cases} 1 & \text{if } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha))[|A|_{w,\gamma} \leq |B|_{w,\gamma}] \\ 0 & \text{if } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha))[|A|_{w,\gamma} = 1 \wedge |B|_{w,\gamma} = 0] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Like the earlier construction with '$\triangleright$', this construction gives rise to a set of final ordinals that include reflection ordinals Δ, and a Fundamental Theorem just like the previous:

Proposition. ***[Fundamental Theorem for L^* (revision-theoretic version).]*** *For any reflection ordinal Δ, any $w \in W$, and any L^+-sentence A,*

(a) $|A|_{w,\Delta} = 1$ if and only if $(\forall\gamma \in FIN)(|A|_{w,\gamma} = 1)$

(b) $|A|_{w,\Delta} = 0$ if and only if $(\forall\gamma \in FIN)(|A|_{w,\gamma} = 0)$.

It can be shown that if the 'True'-free fragment L_0^* is 2-valued, $\to$ and $\Rightarrow$ are each equivalent to the material conditional $\supset$ on L_0^*. (If the 'True'-free fragment L_0^* is 3-valued, as it would be if we were to add $\triangleright$ to the language and used the modified-Burgess-based semantics, then $\Rightarrow$ behaves on it like the Lukasiewicz 3-valued conditional, and $\to$ like a less familiar one.)

As with $\triangleright$, only the valuations at reflection ordinals are relevant to validity: an inference is valid iff in all starting models and all worlds w in them and all reflection Δ, if the premises have value 1 at w and Δ then so does the conclusion.

Alternatively, we could adapt the fixed point semantics, to get a set $\mathbf{R}$ of valuations u assigning values in $\{0, \frac{1}{2}, 1\}$ to each $\to$-conditional at each world, with privileged member v^*. Again, the semantics for non-privileged members of $\mathbf{R}$ is given by a somewhat complicated chain construction analogous to that in Appendix B, but again it very much simplifies for v^*: we get

Proposition. ***[Fundamental Theorem for L^* (fixed point version).]*** *For any $w \in W$, and any L^{*+}-sentence A,*

(a)$|A|_{w,v*} = 1$ if and only if $(\forall u \in \mathbf{R})(|A|_{w,u} = 1)$

(b) $|A|_{w,v*} = 0$ if and only if $(\forall u \in \mathbf{R})(|A|_{w,u} = 0)$.

Only the special v^* is used in the definition of validity.[33]

I note two consequences of the Fundamental Theorems for L^*:

Modus Ponens for $\to$ and $\Rightarrow$: $A, A \to B \vDash B$ (and hence $A, A \Rightarrow B \vDash B$)

Weak Equivalence of $\neg(A \to B)$ and $\neg(A \Rightarrow B)$ to $A \wedge \neg B$: The inference from either $\neg(A \to B)$ or $\neg(A \Rightarrow B)$ to $A \wedge \neg B$ is valid, and so are the reverse inferences.

[33]The difference between the fixed point constructions for $\to$ and for $\triangleright$ comes in the way that chains of valuations generate valuations: instead of the association given in Appendix B, here when Z is a chain of $\to$-valuations we use the much simpler:

$$val[Z](w, A \to B) = \begin{cases} 1 & \text{if } (\exists S \in Z)(\forall u \in S)(|A|_{w,u} = 1 \supset |B|_{w,u} = 1) \\ 0 & \text{if } (\exists S \in Z)(\forall u \in S)(|A|_{w,u} = 1 \wedge |B|_{w,u} = 0) \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

This is basically what's in Section 7 of Field 2014.

Why is the second one called "Weak" Equivalence? Two reasons: (a) While (in the revision version) $|\neg(A \to B)|_{w,\Delta}$ (or $|\neg(A \Rightarrow B)|_{w,\Delta}$) is 1 iff $|A \wedge \neg B|_{w,\Delta} = 1$, there is no analogous claim for 0. (b) Even for 1, the result holds only for reflection Δ, not for all final ordinals. (Similarly in the fixed point case: the equivalence holds only at v^*, not at all valuations in **R**.) A consequence of (b) is that $\neg(A \to B)$ won't in general be intersubstitutable with $A \wedge \neg B$ even in positive contexts, unless those contexts are outside the scope of $\to$'s.

The proofs of Modus Ponens and Weak Equivalence are routine applications of the Fundamental Theorem (for the appropriate construction) together with the evaluation clauses for $\to$. (Here there is no dependence on any Weak Centering assumption since the $\to$ construction operates only within worlds.)

Later I will use the following (stated here for the revision-theoretic construction, but with analogs for fixed point): for all worlds w, and all ordinals α for (L-i) and all reflection ordinals Δ for (L-ii):

(L-i): If $|A \to B|_{w,\alpha} = 1$ then $|B \to C|_{w,\alpha} \leq |A \to C|_{w,\alpha}$;
(L-ii): $|A \to (B \wedge C)|_{w,\Delta} = \min\{|A \to B|_{w,\Delta}, |A \to C|_{w,\Delta}\}$.

The analogs for '$\Rightarrow$' hold as well. Verification of (L-i) is almost trivial. (I'll actually use it only in the case where α is a reflection ordinal, but it holds for all ordinals α.) Part of (L-ii) also generalizes to all ordinals:

(L-iia): If $|A \to B|_{w,\alpha} = 1$ then $|A \to C|_{w,\alpha} \leq |A \to (B \wedge C)|_{w,\alpha}$

(and similarly for $\Rightarrow$), which is likewise easily proved. The remainder of (L-ii) is that when $|A \to (B \wedge C)|_{w,\Delta} = 0$, one of $|A \to B|_{w,\Delta}$ and $|A \to C|_{w,\Delta}$ must be 0. That's so because if $|A \to B|_{w,\Delta}$ and $|A \to C|_{w,\Delta}$ are both > 0 then (by the Fundamental Theorem and the evaluation rules) either there's a final α with $|A|_{w,\alpha} < 1$, or both a final α with $|B|_{w,\alpha} > 0$ and a final β with $|C|_{w,\beta} > 0$; and then by the Fundamental Theorem again, either $|A|_{w,\Delta} < 1$, or both $|B|_{w,\Delta} > 0$ and $|C|_{w,\Delta} > 0$. So $|A \to (B \wedge C)|_{w,\Delta+1} > 0$ and (by the Fundamental Theorem once again) $|A \to (B \wedge C)|_{w,\Delta} > 0$.[34]

6. The two types of conditionals together

So, we know several ways of getting naive truth in a language L with '$\rhd$', and corresponding ways of getting naive truth in a language L^* with '$\to$'. But what we really want is a language L^{**} with both (and with no restrictions on the embedding of either within the scope of the other).

There are three *prima facie* possible ways to proceed.

The *symmetric option* is to give a single construction (revision or fixed point, as one chooses) that evaluates both kinds of conditionals simultaneously: on the revision approach, this would involve, at each stage α, evaluating both $|A \rhd B|_{w,\alpha}$ and $|A \to B|_{w,\alpha}$ on the basis of the various $|A|_{x,\beta}$ and $|B|_{x,\beta}$ for $\beta < \alpha$ (restricting to the case where x is w in the case of $\to$).

The $\rhd$*-first option* is to temporarily hold a valuation v for $\to$ fixed, and use a construction for $\rhd$ on the basis of it. In the case of a revision construction, this would lead, for each choice of v, to a reflection ordinal Ω_v and thus a privileged valuation j^v $(= {j^v}_{\Omega_v})$ for $\rhd$; in the case of a fixed point construction we similarly get a privileged valuation $j{*}^v$. Call this the "inner construction". We then would give

[34] Had we used the valuation rules for the "Official Conditional" of Field 2008, we would only have gotten (L-iia), not (L-ii).

an "outer" construction (again either revision-theoretic or fixed point; and it needn't be the same choice as for the inner) of a valuation for $\rightarrow$, one that looks only at the privileged valuations of $\triangleright$-conditionals constructed in inner constructions from other valuations. For instance, in the case where both the inner and outer constructions are revision-theoretic, we would construct $v_{\alpha+1}$ using valuations of sentences where $\rightarrow$-conditionals are evaluated by v_α and $\triangleright$-conditionals by the corresponding j^{v_α} (and use the same rule for limit ordinals as before), eventuating in a reflection ordinal Δ for the whole construction.

The *$\rightarrow$-first option* is just the reverse. In the case when both inner and outer constructions are revision-theoretic, we temporarily hold fixed a valuation j for $\triangleright$, and use a revision construction for $\rightarrow$ on the basis of it; this leads, for each choice of j, to a reflection ordinal Δ_j and thus a privileged valuation v^j $(= v^j_{\Delta_j})$ for $\rightarrow$. That is the "inner construction". We then would give an "outer" construction of a valuation j for $\triangleright$, where each $j_{\mu+1}$ is determined from an evaluation of sentences that uses j_μ and the corresponding v^{j_μ}, eventuating in a reflection ordinal Ω for the whole construction.

These three choices lead to significantly different results for the joint logic of $\triangleright$ and $\rightarrow$. I think the $\rightarrow$-first option is most natural: very roughly, it involves settling the valuation of $\rightarrow$ at each world before doing the $\triangleright$-construction which relates different worlds. But the ultimate rationale for the $\rightarrow$-first option is that it leads to by far the most plausible and useful laws of restricted quantification.[35] Some of the laws it leads to will be listed in Section 7. Few of them would hold on either the symmetric or $\triangleright$-first options: in the case of the revision construction, that's because on those options, the validity of a sentence of form $A \triangleright B$ (where A and B may contain $\rightarrow$) would require that B has value 1 when A does *at all final ordinals* in the $\rightarrow$-construction, not just at reflection ordinals of the $\rightarrow$-construction. For instance, it's only at reflection ordinals where A and $A \rightarrow \bot$ are prevented from simultaneously having value 1; because of this, the law $[(A \rightarrow B) \wedge A] \triangleright B$ couldn't possibly hold on the symmetric or $\triangleright$-first options, where it does on the $\rightarrow$-first. (Similar remarks hold for the fixed point constructions.) For more remarks related to this, see note 42 below.

Let's recap (or make explicit) how the overall construction goes on the $\rightarrow$-first option. (I'll stick to the case where both the inner and outer constructions are revision-theoretic.) We start with a 2-valued worlds model M_0 for the 'True'-free fragment of L^{**} (whose number-theoretic part is an ω-model in each world, as before). Its ground fragment ${L^{**}}_0$ is to be evaluated either by Burgess 2-valued or variant-Burgess 3-valued semantics. In the former case, '$\rightarrow$' is to be evaluated like '$\supset$' in the ground language. In the latter case, it is to be evaluated in the ground language by the rule that $|A \rightarrow B|$ is 1 whenever $|A| < 1$ and is $|B|$ when $|A| = 1$. (This leads to $\Rightarrow$ being evaluated in the ground language by the 3-valued Lukasiewicz rules: $|A \Rightarrow B|$ is 1 iff $|A| \leq |B|$, 0 iff $|A|$ is 1 and $|B|$ is 0, $\frac{1}{2}$ iff $|A|$ exceeds $|B|$ by $\frac{1}{2}$.) For convenience we expand the language L^{**} by adding names for all objects in the domain U of M_0, getting L^{**+}.

[35]Field 2014 used fixed point constructions rather than revision constructions for inner and outer, but the decision to take the restricted quantifier conditional as inner was the same there as here. (Recall from note 31 the confusing difference in notation: the restricted quantifier conditional there was $\blacktriangleright$, and the $\rightarrow$ there was somewhat in the spirit of the $\triangleright$ here.) The inner construction there was called the "fiber construction", and the outer construction the "base space construction".

Now let T be any function that assigns to every object of the ground model a value in $\{0, \frac{1}{2}, 1\}$, subject to the condition that if an object isn't the Gödel number of a sentence of L^{**}, T assigns it 0. Let j be any function that assigns to every L^{**+}-sentence of form $A \triangleright B$ a value in $\{0, \frac{1}{2}, 1\}$, and v be any function that assigns to every L^{**+}-sentence of form $A \rightarrow B$ a value in $\{0, \frac{1}{2}, 1\}$. We now evaluate every L^{**+}-sentence relative to T, j, and v by essentially the Kleene rules early in Section 4; the only differences are that there is an additional parameter v in all the valuations, and we have an additional trivial clause for v analogous to that for j:

$|A \rightarrow B|_{w,j,v,T} = v(w, A \rightarrow B)$.

Then, keeping j and v fixed, we construct the minimal fixed point T_{min} (which now depends on v as well as on M_0 and j), and abbreviate $|A|_{w,j,v,T_{min}}$ as $|A|_{w,j,v}$.

Next we do the "inner construction": we hold the valuation j for $\triangleright$-sentences fixed, and do a revision construction for valuations v_α of $\rightarrow$-sentences. Adding a subscript j to make explicit the dependence on that $\triangleright$-valuation, the stages are given by:

$$|A \rightarrow B|_{w,j,\alpha} = \begin{cases} 1 & \text{if } (\exists\beta < \alpha)(\forall\gamma \in [\beta, \alpha))[|A|_{w,j,\gamma} = 1 \supset |B|_{w,j,\gamma} = 1] \\ 0 & \text{if } (\exists\beta < \alpha)(\forall\gamma \in [\beta, \alpha))[|A|_{w,j,\gamma} = 1 \wedge |B|_{w,j,\gamma} = 0] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

For each j, we are led to reflection ordinals Δ (which may depend on j as well as on M_0). And the dependence on a j clearly does nothing to block the Fundamental Theorem: we have

Proposition. ***[Fundamental Theorem for*** $\rightarrow$***in*** *L^{**}.] For any j, any j-reflection ordinal Δ, any $w \in W$, and any L^{+}-sentence A,*

(a) $|A|_{w,j,\Delta} = 1$ if and only if $(\forall\gamma \in FIN)(|A|_{w,j,\gamma} = 1)$

and (b) $|A_{w,j,\Delta} = 0$ if and only if $(\forall\gamma \in FIN)(|A|_{w,j,\gamma} = 0)$.

Since there is only one possible value other than 0 and 1, these two clauses imply that each j-reflection ordinal Δ is associated with the same $\rightarrow$-valuation v^{Δ}; we can call this valuation $v(j)$. Since the particular Δ doesn't matter as long as it is a j-reflection ordinal, we can define $|A|_{w,j}$ to be $|A|_{w,j,\Delta}$ where Δ is any j-reflection ordinal.

In short, for each j-valuation for $\triangleright$-sentences, we've assigned a privileged valuation $v(j)$ for $\rightarrow$-sentences. (And a minimal fixed point for truth, based on both.) That's the inner construction.

We now use the privileged $v(j)$'s for each j in constructing a specific j for $\triangleright$-sentences (the "outer construction"). So unlike in the inner construction, we don't need to add a new parameter v for the valuation of the other conditional $\rightarrow$: the clauses for the j_μ that evaluate $\triangleright$-sentences are EXACTLY as in Section 4.

This may seem to simplify matters, but it actually makes them somewhat more complicated: for the v we use is no longer held constant, it varies with the j in the revision process. Because of this, we need to revisit the Fundamental Theorem for $\triangleright$: in particular, the induction on complexity in Stage (2) of the proof in Appendix A. For now we must consider in the induction step not only sentences of form $\neg B$, $B \wedge C$, $\forall x B$ and $\Box A$, but also sentences of form $A \rightarrow B$. And it's unobvious how to carry out the induction step in this case.

Indeed, it's more than unobvious: it can't be done, **the Fundamental Theorem for $\triangleright$ no longer holds without restriction once $\rightarrow$ is added to the**

language. Example: As a preliminary, let $K_\triangleright$ be constructed (by the usual Gödel-Tarski procedure) to be equivalent to $True(\langle K_\triangleright\rangle) \triangleright \neg True(\langle K_\triangleright\rangle)$, and hence given naivety to $K_\triangleright \triangleright \neg K_\triangleright$. On the semantics as given, at each world w for which $W_w \neq \emptyset$ (which includes all those w at which there is at least Weak Centering) and for each stage κ for the outer construction,

(\$): $|K_\triangleright|_{w,\kappa}$ is 1 if κ is odd, 0 if κ is an even successor, and $\frac{1}{2}$ if κ is a limit.

(That's so both for the semantics based on the original Burgess and the one based on the variant.) Now let K^* be $K_\triangleright \rightarrow \neg K_\triangleright$. Since $K_\triangleright$ is equivalent to a $\triangleright$-conditional, its value is held fixed during any $\rightarrow$-construction, so at each w and each stage κ for the outer construction and each stage $\alpha > 0$ for the inner, $|K^*|_{w,\kappa,\alpha}$ is 1 if $|K_\triangleright|_{w,\kappa} < 1$ and is 0 otherwise. So using (\$), when $W_w \neq \emptyset$, $|K^*|_{w,\kappa}$ (i.e. $|K^*|_{w,\kappa,\Delta}$) is 1 when κ is even (including when it is a limit), and 0 otherwise. So for any world, at $\kappa = \Omega$, K^* has value 1, but not at all final κ in the $\triangleright$-construction.[36]

The failure of the Fundamental Theorem for $\triangleright$ is not devastating, for we still get the special case of it for $\triangleright$-conditionals, which is what is needed for many laws, such as Modus Ponens (assuming Weak Centering for $\triangleright$). Indeed, we get more generally that the Fundamental Theorem for $\triangleright$ holds for every sentence A in which all occurrences of 'True' and '$\rightarrow$' are inside the scope of an '$\triangleright$'.

The special case of the Fundamental Theorem for $\triangleright$ is enough to establish that all reflection ordinals in the j_ν construction give rise to the same values for every sentence: for it immediately gives this for every $\triangleright$-conditional, and the generalization to all sentences is immediate by induction.

7. Application to restricted quantification

Here are some highly desirable laws of restricted quantification: it is hard to imagine making serious use of restricted quantification without them, or at least, something very close to them. Indeed, we should expect them to be strong laws in the sense explained in Section 4.1, which guarantees that prefixing any string of $\Box$s and $\Diamond$s to one of them is also to be a law, and that they remain valid however their antecedent is strengthened.[37] (The four with an asterisk are obtained using $\triangleright$-contraposition from their unasterisked counterparts;[38] but since $\triangleright$-contraposition isn't generally valid for variably strict conditionals they need to be stated separately. The ones marked 'b' result from the corresponding ones marked 'a' by a kind of quasi-contraposition which is also not generally valid for variably strict conditionals.) I've written these laws with $\Rightarrow$, reflecting my view that the conditional for restricted quantification is contraposable, but until we get to CQ, every law on the list would remain valid were $\Rightarrow$ to be replaced with $\rightarrow$.

1: $[\forall x(Ax \Rightarrow Bx) \wedge Ay] \triangleright By$ "If all A are B, and y is A, then y is B"
2: $\forall x Bx \triangleright \forall x(Ax \Rightarrow Bx)$ "If everything is B, then all A are B"
2*: $\neg\forall x(Ax \Rightarrow Bx) \triangleright \neg\forall x Bx$ "If not all A are B, then not everything is B"

[36] It won't help to alter the starting point of the $\rightarrow$-construction, e.g. by making conditionals start with value $\frac{1}{2}$ at some worlds but 1 at some and 0 at others. There are several reasons, but the main one is that the evaluation of $K_\triangleright$ would even out by stage ω, so that (\$) would still hold for infinite κ.

[37] Note that though the proof of the latter in note 26 relied on the Fundamental Theorem, it used it only for $\triangleright$-sentences, so it still holds when $\rightarrow$ is in the language.

[38] With double negation laws (and re-lettering) in the case of CQ*.

3a: $\forall x(Ax \Rightarrow Bx) \wedge \forall x(Bx \Rightarrow Cx) \rhd \forall x(Ax \Rightarrow Cx)$
"If all A are B and all B are C then all A are C"
3b: $\forall x(Ax \Rightarrow Bx) \wedge \neg\forall x(Ax \Rightarrow Cx) \rhd \neg\forall x(Bx \Rightarrow Cx)$
"If all A are B and not all A are C then not all B are C"
4a: $\forall x(Ax \Rightarrow Bx) \wedge \forall x(Ax \Rightarrow Cx) \rhd \forall x(Ax \Rightarrow Bx \wedge Cx)$
"If all A are B and all A are C then all A are both B and C"
4b: $\forall x(Ax \Rightarrow Bx) \wedge \neg\forall x(Ax \Rightarrow Bx \wedge Cx) \rhd \neg\forall x(Ax \Rightarrow Cx)$
"If all A are B and not all A are both B and C then not all A are C"
4a*: $\neg\forall x(Ax \Rightarrow Bx \wedge Cx) \rhd \neg\forall x(Ax \Rightarrow Bx) \vee \neg\forall x(Ax \Rightarrow Cx)$
"If not all A are both B and C then either not all A are B or not all A are C"
5: $\neg\forall x(Ax \Rightarrow Bx) \rhd \exists x(Ax \wedge \neg Bx)$
"If not all A are B, then something is both A and not B"
5*: $\forall x(\neg Ax \vee Bx) \rhd \forall x(Ax \Rightarrow Bx)$ "If everything is either not-A or B, then all A are B" / "If nothing is both A and not-B, then all A are B"
6: $\exists x(Ax \wedge \neg Bx) \rhd \neg\forall x(Ax \Rightarrow Bx)$
"If something is both A and not B, then not all A are B"
CQ: $\forall x(Ax \Rightarrow Bx) \rhd \forall x(\neg Bx \Rightarrow \neg Ax)$
"If all A are B then all not-B are not-A".
CQ*: $\neg\forall x(Ax \Rightarrow Bx) \rhd \neg\forall x(\neg Bx \Rightarrow \neg Ax)$
"If not all A are B then not all not-B are not-A".

(There is a bit of redundancy in the list: 2* follows by obvious laws from 5, and 2 from 5*.)

CQ and CQ* strike me as less *obviously* desirable than the earlier members of the list. However, CQ together with 1 and 2 respectively (and double negation laws in the case of 2) yield:

1c: $[\forall x(Cx \Rightarrow Dx) \wedge \neg Dy] \rhd \neg Cy$
"If all C are D, and y is not D, then y is not C"
2c: $\forall x\neg Cx \rhd \forall x(Cx \Rightarrow Dx)$ "If nothing is C, then all C are D"

And these do seem to me obviously desirable; indeed, no less so than the laws 1 and 2 from which they were obtained. It's unobvious how to get a plausible theory that delivers 1c and 2c without delivering CQ (and probably CQ*), which I take to provide support for the latter. Still, someone willing to give up 1c and 2c could use the results of this paper to validate the laws of restricted quantification preceding CQ with a restricted quantifier based on $\rightarrow$ instead of $\Rightarrow$.

Despite the desirability of these laws, it is not entirely easy to give an account of conditionals in naive truth theory that validate them all (even without the modal prefixes). Indeed, prior to Field 2014, no published theory came close. But there are two precursors worth mentioning, Beall *et al* 2006 and Beall 2009. Both are in a paraconsistent framework, which means (given reasonable assumptions that they accept) that they can't accept a restricted-quantifier analog of law 2c, or even of its rule form. For if Cx means $x = x \wedge A$ and Dx means $x = x \wedge B$ then even the rule version of 2c requires that $\neg A$ imply $A \mapsto B$ (where $\mapsto$ is the paraconsistent restricted quantifier conditional); and then Modus Ponens yields Explosion. To deal with this, both precursors propose that the conditional that restricts quantification be non-contraposable,[39] i.e. they disallow even the rule form of CQ for $\mapsto$ (and

[39]Interestingly, they take their main conditional (a relevance conditional, their analog of my $\rhd$) to obey a rule form of contraposition. (Beall 2009 very clearly does; Beall *et al* 2006 is slightly

CQ*, given previous note). Myself, I'm not happy with the loss of 2c; but neither account does well with other laws either.

Beall *et al* 2006 made an important contribution in focusing on the need of a logic of restricted quantification and introducing the idea of using two separate conditionals for it. The paper didn't show, or even claim, that a naive truth theory could be added without triviality to the main logics it considers (those in their Section 6); but their discussion is explicitly motivated by the hope/belief that this is so. (One of the authors explicitly stated several years later that the question of non-triviality was open: see Beall 2009, p. 121.) Putting any worries about lack of non-triviality proof aside, the main issue is over the laws. The good news is that their framework validates their analogues of laws 2 and 4a (taking the analogues to have their noncontraposable $\mapsto$ in place of my contraposable $\Rightarrow$, as well as their relevance conditional in place of my $\triangleright$); hence also 2* and 4a*, assuming the interpretation in note 39. The bad news is that it doesn't validate any of the others (though it does validate rule forms of some of them). Also, the validation of 2 and 2* depends very directly on their assumption of the validity of

(?): $A \triangleright B \models A \mapsto B$.

And (?) immediately rules out the analog of my law 1 (when naive truth, Modus Ponens for $\mapsto$, and reasonable quantifier laws are present). The reason is that given reasonable quantifier laws, law 1 requires $[(A \mapsto B) \wedge A] \triangleright B$; and then (?) delivers

Pseodo Modus Ponens: $[(A \mapsto B) \wedge A] \mapsto B$.

And it's well-known that this is inconsistent with genuine Modus Ponens for $\mapsto$ (i.e. $(A \mapsto B) \wedge A \models B$) in a naive theory (assuming the standard structural rules for validity mentioned in note 3).[40] The centrality of (?) to the derivation of law 2 suggests that no simple modification of the account is likely to yield laws 1 and 2 together.

The second precursor is Beall 2009 (pp. 119-226). It also used two separate conditionals for the logic of restricted quantification. It suggests three different options for the logic, and unlike Beall *et al* 2006, shows each to be compatible with naive truth. All of them validate (?), so again it is immediate that law 1 can't be satisfied. The situation for laws is slightly worse than Beall *et al* 2006. Beall's first two options validate only 4a and 4a* from the list (though the weaker rule forms of some of the others are validated). His third option validates only 2 and 2*; indeed, its method of achieving 2 and 2* causes it to violate even the rule form of 4a.

Without going into detail, the main problem in both Beall *et al* 2006 and Beall 2009 arises because (a) a certain kind of "abnormal" worlds are essential to these accounts (unlike the present account, where they are optional); (b) at these worlds, both conditionals are very badly behaved; and (c) the validity of $X \triangleright Y$ (using my notation for their relevance conditional) requires that it be true at all normal worlds, which in turn requires that at all worlds including abnormal ones, Y is true when X is. Collectively these make it very hard for reasonable $\triangleright$-statements

---

equivocal: see p. 595 middle.) I take this to mean that their main conditional isn't a good candidate for an account of the ordinary indicative conditional: see the Trump example in Section 2.

[40] In the logic I've been advocating (with Weak Centering assumed so as to get Modus Ponens), we do have

$$C \vee \neg C, C \triangleright B \models C \Rightarrow B;$$

but the need for the excluded middle premise is sufficient to prevent the paradox.

with $\mapsto$-conditionals in their antecedents or consequents to come out valid. (An additional problem arises because of the way that these accounts handles negation, via a shift in worlds: this immediately rules out laws like 3b and 4b.)

Field 2014 used a very different framework, and did manage to validate the entire list; but the semantics it employed for $\triangleright$ seemed *ad hoc*. (That paper did note some commonalities between its $\triangleright$ and the ordinary indicative conditional, but also pointed out that the conditional reduced to the material conditional rather than the indicative conditional in 'True'-free contexts.)[41]

But I now note that the entire list is also validated on the semantics of the present paper, with its independently motivated $\triangleright$. (We also get Modus Ponens for $\triangleright$, if we insist on Weak Centering at normal worlds in the base model, as I think we clearly should.)

The real work in establishing the laws on the list has nothing to do with the quantifiers, it's all in the relation among conditionals. The laws we need are the results of prefixing the following with strings of $\Box$s and $\Diamond$s:

I: $[(A \Rightarrow B) \wedge A] \triangleright B$ (for 1)
IIIa: $(A \Rightarrow B) \wedge (B \Rightarrow C) \triangleright (A \Rightarrow C)$ (for 3a)
IIIb: $(A \Rightarrow B) \wedge \neg(A \Rightarrow C) \triangleright \neg(B \Rightarrow C)$ (for 3b)
IVa: $(A \Rightarrow B) \wedge (A \Rightarrow C) \triangleright (A \Rightarrow B \wedge C)$ (for 4a)
IVb: $(A \Rightarrow B) \wedge \neg(A \Rightarrow B \wedge C) \triangleright \neg(A \Rightarrow C)$ (for 4b)
IVa*: $\neg(A \Rightarrow B \wedge C) \triangleright \neg(A \Rightarrow B) \vee \neg(A \Rightarrow C)$ (for 4a*)
V: $\neg(A \Rightarrow B) \triangleleft\triangleright (A \wedge \neg B)$ (for 5, 2* and 6)
V*: $(\neg A \vee B) \triangleright (A \Rightarrow B)$ (for 5* and 2)
C: $(A \Rightarrow B) \triangleleft\triangleright (\neg B \Rightarrow \neg A)$ (for CQ)
C*: $\neg(A \Rightarrow B) \triangleleft\triangleright \neg(\neg B \Rightarrow \neg A)$ (for CQ*)

C and C* are of course entirely trivial given the definition of $\Rightarrow$ in terms of $\rightarrow$. For most of the others, the proof is almost immediate from what has already been said, especially at the end of Section 5. (The analogs of these latter laws for $\rightarrow$ hold equally.) For note that to establish that a claim of form $P(X \triangleright Y)$ is valid, where P is any string of $\Box$s and $\Diamond$s, it suffices to show (in the revision-theoretic version; but the fixed point is analogous) that for all worlds w and all final κ of the $\triangleright$-construction, if $|X|_{w,\kappa} = 1$ then $|Y|_{w,\kappa} = 1$. In other words, that for all w and κ, and all κ-reflection ordinals Δ_κ of the $\rightarrow$-construction, if $|X|_{w,\kappa,\Delta_\kappa} = 1$ then $|Y|_{w,\kappa,\Delta_\kappa} = 1$. Given this, the proof of I is immediate from "Modus Ponens for $\rightarrow$ and $\Rightarrow$", and V from "Weak equivalence of $\neg(A \rightarrow B)$ and $\neg(A \Rightarrow B)$ to $A \wedge \neg B$". And IIIa and IIIb follow from the special case of (L-i) (end of Section 5) where α is Δ, and IVa, IVb and IVa* from (L-ii). As for V*, if $|\neg A \vee B|_{w,\kappa,\Delta_\kappa} = 1$ then either $|A|_{w,\kappa,\Delta_\kappa} = 0$ or $|B|_{w,\kappa,\Delta_\kappa} = 1$, and so by the Fundamental Theorem for $\rightarrow$, either for all κ-final α, $|A|_{w,\kappa,\alpha} = 0$ or else for all κ-final α, $|B|_{w,\kappa,\alpha} = 1$; in either

[41] Despite its reducing to the material conditional, we can in retrospect see the conditional of Field 2014 as pretty much a degenerate case of the indicative conditional of the present paper. For the construction there started from a classical first order model, which can be seen as a degenerate Burgess model with only one world, weakly centered (which in the one-world case means simply "accessible from itself"). In that degenerate case, '$\triangleright$' obviously coincides with the material conditional in the ground model. (The conditional there still differed in a small respect from the degenerate case of the current construction: it utilized what I there called "dynamic Kripke constructions". I have dropped them here since they don't yield the results that we want once we clearly focus on extending the *ordinary indicative* conditional to a language with 'True'.)

case, for all κ-final α, $|A|_{w,\kappa,\alpha} \leq |B|_{w,\kappa,\alpha}$. From this it clearly follows that for all final α $|A \Rightarrow B|_{w,\kappa,\alpha} = 1$ and hence in particular that $|A \Rightarrow B|_{w,\kappa,\Delta_\kappa} = 1$.

This only scratches the surface of the logic of the system,[42] but it is not my purpose here to explore it at all systematically: my purpose was simply to show that it does easily lead to obvious laws of restricted quantification, which other approaches to conditionals in naive truth theory (other than the *ad hoc* one of Field 2014) haven't come close to meeting. And I think that by basing the laws on an independently motivated account of indicative conditionals, the resulting theory is quite natural.

In particular, it's worth emphasizing that the use of two distinct conditional operators (which is essential for the compatibility of the logic with naive truth, since if $\rightarrow$ and $\triangleright$ were identified then we'd have the disastrous (?)) was independently motivated: as I argued at the beginning of Section 5, we can see independently of the laws recently listed that the indicative conditional and the conditional for restricted quantification must be different.

Thanks: Harvey Lederman, Graham Priest and two anonymous referees made comments that have led to significant improvements.

References

[1] Beall, JC 2009. *Spandrels of Truth*. Oxford University Press.
[2] Beall, JC, Ross Brady, Allen Hazen, Graham Priest and Greg Restall 2006. "Relevant Restricted Quantification." *Journal of Philosophical Logic* 35: 587-598.
[3] Burgess, John 1981. "Quick Completeness Proofs for some Logics of Conditionals." *Notre Dame Journal of Formal Logic* 22: 76-84.
[4] Field, Hartry 2008. *Saving Truth from Paradox*. Oxford University Press.
[5] Field, Hartry 2014. "Naive Truth and Restricted quantification: Saving Truth a Whole Lot Better." *Review of Symbolic Logic* 7:147-91.
[6] Gupta, Anil, and Nuel Belnap 1993. *The Revision Theory of Truth*. MIT Press.
[7] Kripke, Saul 1975. "Outline of a Theory of Truth." *Journal of Philosophy* 72: 690-716.
[8] Lewis, David 1974. *Counterfactuals*. Harvard University Press.
[9] McGee, Vann 1985. "A Counterexample to Modus Ponens". *Journal of Philosophy* 82: 462-71.
[10] Pollock, John 1976. "The 'Possible Worlds' Analysis of Counterfactuals." *Philosophical Studies* 29: 469-76.

[42]The reader will note that the schemas I've listed and proved are ones where there are no occurrences of $\triangleright$ inside the scope of an $\rightarrow$ (or an $\Rightarrow$). This is no accident: the $\rightarrow$-first construction makes it much easier for a schema in which $\rightarrow$ is in the scope of $\triangleright$ to be valid than for one where $\triangleright$ is in the scope of $\rightarrow$ to be valid. I think that schemas of the latter sort tend to be far less important than the former (recall the frequently-voiced claim that embeddings of indicative conditionals in the scope of other operators are hard to interpret); that is the main reason I went for an $\rightarrow$-first option. (I have however made no prohibitions on the well-formedness of embeddings of $\triangleright$ inside the scope of $\rightarrow$; and with any valid schema such as those listed, there are instances of the schema with arbitrarily complex chains of embeddings of $\triangleright$ and $\rightarrow$.)

Despite what I've just said, there are important laws that depend on the embedding of '$\triangleright$' in the scope of '$\rightarrow$', but these are mostly meta-rules, whose legitimacy is not blocked by the $\rightarrow$-first option. A typical example is the meta-rule stated in note 28, whose proof is routine.

[11] Restall, Greg 2007. "Curry's Revenge: The Costs of Non-Classical Solutions to the Paradoxes of Self-Reference." In JC Beall, ed., *Revenge of the Liar*, Oxford University Press.
[12] Stalnaker, Robert 1968. "A Theory of Conditionals." In N. Rescher, ed., *Studies in Logical Theory*, Blackwell Publishers.
[13] Yablo, Stephen 2003. "New Grounds for Naive Truth Theory." In JC Beall ed., *Liars and Heaps*, Oxford University Press.

Appendix A: Proof of Fundamental Theorem for L (revision-theoretic version)

Since $\Omega \in FIN$, the right to left of (a) and (b) in the Theorem are trivial. Contraposing the left to right and making the Kripke-stages σ explicit, what we need to establish is that for any reflection ordinal Ω and any L^+-sentence A:

(a*) $(\forall w \in W)$[if $(\exists \nu \in FIN)(|A|_{w,\nu} < 1)$ then $\forall\sigma(|A|_{w,\Omega,\sigma} < 1)$], and

(b*) $(\forall w \in W)$[if $(\exists \nu \in FIN)(|A|_{w,\nu} > 0)$ then $\forall\sigma(|A|_{w,\Omega,\sigma} > 0)$].

We establish (a*) and (b*) in three steps:

(1) In the special case when A is a conditional $B \rhd C$, the value of σ makes no difference, and by the fact that Ω is a limit ordinal and the evaluation rules for conditionals are continuous with respect to 1 and 0 at limits, the claims are just:

(a*-s) $(\forall w \in W)$[if $(\exists \nu \in FIN)(j_{w,\nu}(B \rhd C) < 1)$ then $(\forall \mu < \Omega)(\exists \nu \in [\mu, \Omega))[j|_{w,\nu}(B \rhd C) < 1)$];

(b*-s) $(\forall w \in W)$[if $(\exists \nu \in FIN)(j_{w,\nu}(B \rhd C) > 0)$ then $(\forall \mu < \Omega)(\exists \nu \in [\mu, \Omega))[j|_{w,\nu}(B \rhd C) > 0)$].

But by (ii) of [Gupta-Belnap] these are trivial.

(2) Given (1), we can show (for any σ) that if (a*) and (b*) hold for the special case where A is of form '$True(c)$' when c denotes the Gödel number of a sentence, then they hold for all L^+-sentences A. This is a routine induction on complexity, counting $\rhd$-sentences as of complexity 0 for the purposes of the induction: the claim is trivial for all other atomic sentences since they keep the same value at every revision-stage ν, and the induction step for sentences $\neg B$, $B \wedge C$, $\forall x B$ and $\Box B$ is easy. For instance for $\Box$: (a) Suppose that for some world w, $(\exists \nu \in FIN)(|\Box B|_{w,\nu} < 1)$. Then $(\exists \nu \in FIN)(\exists y \in W_w)(|B|_{y,\nu} < 1)$; reversing the quantifier order and applying the induction hypothesis, we get that for some $y \in W_w$, $|B|_{y,\Omega,\sigma} < 1$ (for any σ), and so $|\Box B|_{w,\Omega,\sigma} < 1$. (b) Suppose that for some world w, $(\exists \nu \in FIN)(|\Box B||_{w,\nu} > 0)$. Then $(\exists \nu \in FIN)(\forall y \in W_w)(|B||_{y,\nu} > 0)$; so certainly for all y in W_w, $(\exists \nu \in FIN)(|B||_{y,\nu} > 0)$, and by the induction hypothesis for all y in W_w, $|B|_{y,\Omega,\sigma} > 0$ (for any σ); so $|\Box B|_{w,\Omega,\sigma} > 0$ for any σ.

(3) It remains only to show that for all Kripke-stages σ and all c that denote Gödel numbers of sentences, (a*) and (b*) hold for sentences of form '$True(c)$'. But this is trivial when $\sigma = 0$, since $|True(c)|_{w,\Omega,0}$ is always $\frac{1}{2}$. We now show that if it holds for $\sigma = \tau$ then it holds for $\sigma = \tau + 1$. Suppose c denotes B. Then by the assumption about τ and the result (2), we get

$(\forall w \in W)$[if $(\exists \nu \in FIN)(|B|_{w,\nu} < 1)$ then $|B|_{w,\Omega,\tau} < 1$]

and the analog with '> 0' instead of '< 1'; which by the transparency of the j_ν-valuations and the Kripke construction gives

$(\forall w \in W)$[if $(\exists \nu \in FIN)(|True(c)|_{w,\nu} < 1)$ then $|B|_{w,\Omega,\tau} < 1$]

and its analog. But by the valuation rules, $|B|_{w,\Omega,\tau}$ is the same as $|True(c)|_{w,\Omega,\tau+1}$, so the result is established. The case where σ is a limit ordinal is trivial: no sentence of form '$True(c)$' first passes from $\frac{1}{2}$ to another value at a limit stage of the Kripke construction.

Appendix B: The fixed point construction for L

Again, a valuation function is a function that assigns to each world and L^+-conditional a value in $\{0, \frac{1}{2}, 1\}$.

Let a *chain* be a set P of nonempty sets of transparent valuation functions, meeting the condition that if $S_1, S_2 \in P$ then either $S_1 \subseteq S_2$ or $S_2 \subseteq S_1$.

Given a chain P, define a valuation function $val[P]$ ("the valuation function generated by P") as:

$$val[P](w, A \triangleright B) \text{ is } \begin{cases} 1 & \text{if } (\exists S \in P)(\forall j \in S)(\forall x \in W_w)[|A|_{x,j} = 1 \supset (\exists y \leq_w x) \\ & [|A|_{y,j} = 1 \wedge (\forall z \leq_w y)(|A|_{z,j} = 1 \supset |B|_{z,j} = 1)]] \\ 0 & \text{if } (\exists S \in P)(\forall j \in S)[(\forall x \in W_w)[|A|_{x,j} = 1 \supset (\exists y \leq_w x) \\ & [|A|_{y,j} = 1 \wedge (\forall z \leq_w y)(|A|_{z,j} = 1 \supset |B|_{z,j} = 0)] \wedge (\exists x \in W_w)(|A|_{x,j} = 1)]] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

(This is for the semantics based on modified Burgess; for that based on original Burgess, the modification of the 0 clause is obvious.) Clearly each $val[P]$ is transparent, given that members of $\cup P$ are.

Let $P_1 \leq P_2$ mean that every member of P_1 has a subset that's a member of P_2. (Having small members makes a chain bigger.) Chains that are smaller in this ordering generate weaker valuation functions: if $P_1 \leq P_2$ then for all w, $val_w[P] \leq_K val_w[P]$. (That's simply because the 1 clause and 0 clause both have form "$(\exists S \in P)(\forall j \in S)$...".)

Define a sequence J_μ of sets of transparent valuation functions:

$$J_\mu = \{val[P] : P \text{ is a chain and } (\forall \beta < \mu)(\exists S \in P)(S \subseteq J_\beta)\}.$$

For $\mu > 0$ an equivalent and perhaps more intuitive definition of J_μ is: $\{val[P] : P \text{ is a non-empty chain and } (\forall \beta < \mu)(\forall S \in P)(S \subseteq J_\beta)\}$. This is more restrictive about the chains, but it's easy to see that any valuation generated by one of the chains in the original is generated also by one of the more restrictive ones.

If $\mu < \nu$, $J_\nu \subseteq J_\mu$, so obviously we eventually reach a fixed point $\mathbf{J}$. That would be uninteresting if $\mathbf{J}$ were empty, but it can be shown (following the model of Field 2014, section 7) that $\mathbf{J} \neq \emptyset$. So letting $\mathbf{P}$ be the set of $\mathbf{J}$-chains (chains whose members are all subsets of $\mathbf{J}$) we'll have

(FP): $\mathbf{J} = \{val[P] : P \in \mathbf{P}\}$.

This sets up a one-many correspondence between the j in $\mathbf{J}$ and the P in $\mathbf{P}$. (The members of $\mathbf{J}$ are the analogs in this construction of the valuation functions associated with ordinals in FIN in the revision construction.)

The $\leq$-minimal chain is $\{\mathbf{J}\}$; let j^* be the valuation it generates, i.e. $val[\{\mathbf{J}\}]$. This is the analog, in the fixed point construction, of the valuation function at reflection ordinals. We have

$$|A \triangleright B|_{w,j^*} = \begin{cases} 1 & \text{if } (\forall j \in \mathbf{J})(\forall x \in W_w)[|A|_{x,j} = 1 \supset (\exists y \leq_w x) \\ & [|A|_{y,j} = 1 \wedge (\forall z \leq_w y)(|A|_{z,j} = 1 \supset |B|_{z,j} = 1)]] \\ 0 & \text{if } (\forall j \in \mathbf{J})[(\forall x \in W_w)[|A|_{x,j} = 1 \supset (\exists y \leq_w x) \\ & [|A|_{y,j} = 1 \wedge (\forall z \leq_w y)(|A|_{z,j} = 1 \supset |B|_{z,j} = 0)]] \wedge (\exists x \in W_w)(|A|_{x,j} = 1)] \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

In this case the Fundamental Theorem, as stated in the text, concerns the special nature of the valuation function at j^*. Its proof and the proof of the fixed point result are a simple adaptation of that in Section 7 of Field 2014.