Journal of Philosophical Logic
https://doi.org/10.1007/s10992-024-09749-4

Naïve Truth and the Evidential Conditional

Andrea Iacona[1] · Lorenzo Rossi[1]

Received: 25 November 2023 / Accepted: 15 December 2023

Abstract

This paper develops the idea that valid arguments are equivalent to true conditionals by combining Kripke's theory of truth with the evidential account of conditionals offered by Crupi and Iacona. As will be shown, in a first-order language that contains a naïve truth predicate and a suitable conditional, one can define a validity predicate in accordance with the thesis that the inference from a conjunction of premises to a conclusion is valid when the corresponding conditional is true. The validity predicate so defined significantly increases our expressive resources and provides a coherent formal treatment of paradoxical arguments.

Keyword Stoic thesis · Naïve truth · Evidential conditional validity · Paradox

1 Introduction

According to a view that goes back to the Stoics — call it the *Stoic Thesis* — an argument is valid when the conditional formed by the conjunction of its premises as antecedent and its conclusion as consequent is true.[1] Although this view is not very widespread among contemporary logicians, the connection it suggests between arguments and conditionals is rather intriguing. As Iacona has argued, the Stoic Thesis is appreciably more credible than is usually believed: as long as validity is construed in a fairly broad sense — which is not limited to deductive reasoning — and it is assumed that a conditional holds when its antecedent supports its consequent, the equivalence between valid arguments and true conditionals gains plausibility. More precisely, if our language includes the symbol $\rhd$ to represent a conditional so understood, φ is a

[1] The label 'Stoic Thesis' comes from [21]. The reference is to Sextus Empiricus, *Against the Logicians*, II, 417.

✉ Andrea Iacona
andrea.iacona@unito.it

Lorenzo Rossi
lo.rossi@unito.it

1 Center for Logic, Language and Cognition, Department of Philosophy and Education, University of Turin, Turin, Italy

Published online: 01 March 2024

conjunction of formulae, and ψ is a formula, the validity of the inference from φ to ψ can coherently be defined in the metalanguage in terms of the truth of $\varphi \rhd \psi$.[2]

This paper shows how the Stoic Thesis can be accommodated in a consistent formal theory. Instead of simply providing a metalinguistic definition of validity for a language endowed with the symbol $\rhd$, we will use a language which contains, in addition to $\rhd$, a truth predicate and a definable validity predicate, so as to give formal expression to the claim that the inference from φ to ψ is valid just in case $\varphi \rhd \psi$ is true. This equivalence in the object language — which we call the *Formalized Stoic Thesis* — has at least two interesting implications. First, it significantly increases our expressive resources, enabling us to formalize ascriptions of validity to particular arguments or classes of arguments. Second, it provides a coherent formal treatment of paradoxical arguments.

In order to obtain the Formalized Stoic Thesis, we will combine two theories that have been developed for independent reasons. The first is Kripke's theory of truth.[3] As is well known, Kripke suggested a way of formally constructing a partial extension for a truth predicate Tr while retaining *naïveté*, the principle according to which any sentence φ is equivalent to its truth predication $\mathsf{Tr}\ulcorner\varphi\urcorner$, where $\ulcorner\varphi\urcorner$ is a closed term that denotes φ. Such extension is partial because there are sentences — e.g. the Liar sentence λ, which is equivalent to $\neg\mathsf{Tr}\ulcorner\lambda\urcorner$ — such that neither they nor their negation belong to it. Kripke constructed the extension of Tr in stages indexed by ordinals. At stage 0, the extension is empty. At stage 1, the extension contains the truth-free literals that are satisfied in the base model. Once we have stage 1, we can build stage 2, and so on. More generally, the extension of Tr at stage $\alpha + 1$ contains the result of applying a valuation schema to sentences in the extension at stage α. The process of building more and more extensions for Tr eventually stabilizes because it reaches a *fixed point*: there are ordinals δ such that the extension of Tr at δ is identical to the extension of Tr at $\delta + 1$. The stages corresponding to such ordinals provide the proposed interpretation of Tr.

The second theory is the evidential account of conditionals suggested by Crupi and Iacona. On this account, $\varphi \rhd \psi$ is true just in case φ and $\neg\psi$ are incompatible, where incompatibility is defined in modal terms by assuming a set of possible worlds and a relation of comparative distance between them. Crupi and Iacona regard the resulting definition as a reasonably close approximation to the intuition that a conditional holds when its antecedent supports its consequent: to say that $\varphi \rhd \psi$ is true is to say that φ provides a reason for ψ, or equivalently that ψ can be inferred from φ.[4]

The theory of validity that will be outlined combines the two theories just considered as it employs a language which contains both the predicate Tr as interpreted by Kripke and the symbol $\rhd$ as understood by Crupi and Iacona. As is reasonable to expect, this combination requires adjustments on both sides. On the one hand, since Tr is a partial predicate framed in terms of a three-valued logic, we need to adjust the definition of $\rhd$ to such a logic. On the other hand, since $\rhd$ is spelled out in modal terms, this

2 Iacona [21].

3 This theory is outlined in [24].

4 The original formulation of the view is in [4]. Here we will follow the more recent version provided in [6].

requires a semantics in which Kripke's construction is generalized so as to obtain a set of possible worlds.

The structure of the paper is as follows. Section 2 presents the evidential account of conditionals. Section 3 introduces the language and provides some initial definitions. Sections 4 and 5 articulate the Kripkean modal construction that yields a model for the language. Sections 6 and 7 complete the semantics by providing suitable definitions of truth and logical consequence. Section 8 states the Formalized Stoic Thesis and discusses some of its implications. Section 9 adds some conclusive remarks and sketches some prospects for further work.

2 The Evidential Account of Conditionals

On the evidential account of conditionals, $\varphi \rhd \psi$ is true just in case φ and $\neg\psi$ are incompatible in the following sense: either there are no worlds in which φ and $\neg\psi$ are both true, because φ and $\neg\psi$ exclude each other, or there are worlds in which φ and $\neg\psi$ are both true, but such worlds are comparatively remote. The expression 'because φ and $\neg\psi$ exclude each other' in the first disjunct is rendered as a conjunction of two conditions: (a) φ is true in some worlds, (b) ψ is false in some worlds. The second disjunct is spelled out as a conjunction of three conditions: (c) φ and ψ have the same value at least in some of the closest worlds, (d) the worlds in which φ and $\neg\psi$ are both true are more distant than those in which φ is true and $\neg\psi$ is false, and (e) the worlds in which φ and $\neg\psi$ are both true are more distant than those in which $\neg\psi$ is true and φ is false. More precisely, the truth conditions of $\varphi \rhd \psi$ relative to a world w are stated as follows:

Definition 2.1 $\varphi \rhd \psi$ is true in w iff either the conjunction of φ and $\neg\psi$ is false in every world and the following conditions jointly hold:

(a) φ is true in some w'
(b) ψ is false in some w'

or the conjunction of φ and $\neg\psi$ is true in some world and the following conditions jointly hold:

(c) φ and ψ are either both true or both false in some w' among the closest worlds;
(d) for every w' such that φ and $\neg\psi$ are both true in w', some strictly closer w'' is such that φ and ψ are both true in w'';
(e) for every w' such that φ and $\neg\psi$ are both true in w', some strictly closer w'' is such that φ and ψ are both false in w''.

When the first disjunct holds, φ and $\neg\psi$ are *absolutely incompatible*, in the sense that their joint truth is impossible. (a) and (b) guarantee that this impossibility depends on the relation between φ and $\neg\psi$, as they rule out that it holds merely in virtue of the impossibility of φ or the necessity of ψ. When the second disjunct holds — that is, (c)-(e) are satisfied — φ and $\neg\psi$ are *relatively incompatible*, in that their combination is comparatively remote. (c) requires that φ and ψ have the same value — hence φ and $\neg\psi$ have different values — in some of the closest worlds. One way to make sense of this condition is the following: if φ and $\neg\psi$ are relatively incompatible, meaning that their combination is a remote possibility, φ and ψ must be relatively compatible, meaning that their combination is a near possibility, so it is reasonable to rule out that φ and ψ have different values in all the closest worlds. (d) expresses the Ramsey Test as understood in the Stalnaker-Lewis account of conditionals, for it implies that $\neg\psi$ is false in the closest worlds in which φ is true.[5] Note that, given (d), the only interesting case ruled out by (c) is that in which φ is false and ψ is true in the closest worlds. (e) reverses the Ramsey Test, as it implies that φ is false in the closest worlds in which $\neg\psi$ is true, so the incompatibility between φ and $\neg\psi$ turns out to be symmetric.

Crupi and Iacona call *Chrysippus Test* the whole disjunction, and argue that Definition 2.1 provides a plausible analysis of the idea that φ supports ψ. If the first disjunct holds — that is, φ and $\neg\psi$ are absolutely incompatible — φ provides a conclusive reason for accepting ψ. If the second disjunct holds — that is, φ and $\neg\psi$ are relatively incompatible — φ provides a defeasible reason for accepting ψ.[6]

When φ and $\neg\psi$ are not incompatible in the sense just explained, $\varphi \rhd \psi$ is false, which means that φ does not support ψ. The falsity conditions of $\varphi \rhd \psi$ relative to a world w are stated as follows:

Definition 2.2 $\varphi \rhd \psi$ is false in w iff either the conjunction of φ and $\neg\psi$ is false in every world and at least one of the following conditions holds:

(-a) φ is false in every w';
(-b) ψ is true in every w';

or the conjunction of φ and $\neg\psi$ is true in some world and at least one of the following conditions holds:

(-c) for every w' among the closest worlds, either φ is true in w' and ψ is false in w', or φ is false in w' and ψ is true in w';
(-d) some w' is such that φ and $\neg\psi$ are both true in w', and no strictly closer w'' is such that φ and ψ are both true in w'';
(-e) some w' is such that φ and $\neg\psi$ are both true in w', and no strictly closer w'' is such that φ and ψ are both false in w''.

In the classical framework adopted by Crupi and Iacona, $\varphi \rhd \psi$ turns out to be either true or false in every world, so there is no need to provide distinct definitions for truth conditions and falsity conditions. However, the distinction is necessary here because we want to extend the evidential account to a three-valued semantics that preserves the original assignments of truth and falsity.[7]

Although the evidential account is not the only account of conditionals that can be combined with a Kripkean theory of truth, its distinctive logical profile makes it particularly suited to substantiate the Stoic Thesis, which is our ultimate goal. To illustrate this point, we will explain how $\rhd$ behaves with respect to some well known

[5] Stalnaker [36], Lewis [25].

[6] Crupi and Iacona [6] provides a detailed explanation of Definition 2.1, and compares it with the original definition provided in [4].

[7] The original valuation clauses for the evidential conditional suggested by Crupi and Iacona are defined for propositional formulae, while it is part of our project to show that they apply equally well to the classical fragment of a first-order language endowed with a naïve truth predicate.

principles of conditional logic. Arguably, the logical properties of $\rhd$ listed below are desirable on the assumption that $\rhd$ represents a relation of support.

Let us start with three principles that hold for $\rhd$. The first, *Material Implication*, says that $\varphi \rhd \psi$ entails $\varphi \supset \psi$, where the latter is defined in the usual way as $\neg(\varphi \wedge \neg\psi)$. This principle shows that the evidential conditional is stronger than the material conditional, as is reasonable to expect. The second, *AND*, says that $\varphi \rhd \psi$ and $\varphi \rhd \chi$ jointly entail $\varphi \rhd (\psi \wedge \chi)$. Here the rationale is that when φ supports ψ and χ taken separately, it also supports their conjunction. The third, *Contraposition*, says that $\varphi \rhd \psi$ entails $\neg\psi \rhd \neg\varphi$. This principle, which characterizes the evidential account as distinct from other accounts of conditionals, holds in virtue of the symmetry of incompatibility: whenever φ is incompatible with $\neg\psi$, $\neg\psi$ is incompatible with φ, which means that whenever φ supports ψ, $\neg\psi$ supports $\neg\varphi$.[8]

Now we will state three principles that do not hold for $\rhd$. The first, *Monotonicity*, says that $\varphi \rhd \chi$ entails $(\varphi \wedge \psi) \rhd \chi$ for any ψ. This principle characterizes conclusive reasoning as distinct from defeasible reasoning, so it should fail in any logic that intends to leave room for the latter. The second, *Right Weakening*, says that $\varphi \rhd \psi$ entails $\varphi \rhd \chi$ whenever χ logically follows from ψ. Arguably, this principle should fail as well, for a defeasible evidential connection between φ and ψ can be weakened or lost in the step from ψ to χ. The third, *Conjunctive Sufficiency*, says that $\varphi \wedge \psi$ entails $\varphi \rhd \psi$. This principle is clearly dubious on the intended interpretation of $\rhd$, because it cannot be the case that φ supports ψ just in virtue of the fact that φ and ψ both hold.

The list of logical properties just provided shows some crucial differences between the evidential account and other extant accounts of conditionals. For example, the evidential conditional differs both from the classical strict conditional, which preserves Monotonicity and Right Weakening, and from the Stalnaker-Lewis conditional, which invalidates Contraposition but preserves Right Weakening and Conjunctive Sufficiency.[9] As explained above, both conditionals are inadequate to capture the notion of support: one is too strong, the other is too weak. Another example is Rott's difference-making account of conditionals, which resembles the evidential account as far as Material Implication, AND, Monotonicity, Right Weakening, and Conjunctive Sufficiency are concerned, although it does not preserve Contraposition.[10] The difference-making conditional is definitely better than the strict conditional and the Stalnaker-Lewis conditional as an alternative to the evidential conditional for the purpose of formalizing the Stoic Thesis.

We conclude this section with a remark on *Identity*, the principle according to which $\varphi \rhd \varphi$ is a logical truth. In the formulation of the evidential account adopted here, which differs from earlier formulations provided by Crupi and Iacona, this principle does not hold.[11] By conditions (a) and (b) in the first disjunct of Definition 2.1, $\varphi \rhd \psi$ is not

[8] In the system EC offered in [32], based on the original classical version of the evidential account, Material Implication, AND, and Contraposition feature as axioms. [4] argues for Contraposition and discuss some of its implications.

[9] Crupi and Iacona [5] provides a detailed comparative study of these three conditionals in a probabilistic framework.

[10] Rott [33, 34].

[11] Here we follow [6].

true when φ is impossible or ψ is necessary. For example, the following sentences turn out to be false:

(1) If $0 = 1$, snow is white
(2) If snow is white, $0 = 0$

As explained above, the idea behind (a) and (b) is that the incompatibility between φ and $\neg\psi$ must be relational, that is, it cannot simply depend on φ being impossible or ψ being necessary. So we get that an impossible truth supports nothing, and that nothing supports a necessary truth, which is quite plausible. When $\varphi = \psi$, this property of $\rhd$ generates counterexamples to Identity, as in the following cases:

(3) If $0 = 1, 0 = 1$
(4) If $0 = 0, 0 = 0$

In order to avoid such counterexamples, one would have to define absolute incompatibility without conditions (a) and (b), thereby treating conditionals with impossible antecedents or necessary consequents as cases of vacuous truth.[12]

3 The Language $\mathcal{L}_{\mathsf{t},\rhd}$

The presentation of our theory starts from the language. Since we want our model-theoretic construction to be applicable as widely as possible, we will provide a set of general conditions on languages rather than specifying a single language.

Definition 3.1 Let $\mathcal{L}_{\mathsf{t},\rhd}$ be a first-order language with the following properties:

1. $\mathcal{L}_{\mathsf{t},\rhd}$ includes $\neg, \wedge, \forall$ as logical constants;
2. $\mathcal{L}_{\mathsf{t},\rhd}$ includes the unary predicate Tr and the binary connective $\rhd$;
3. for every formula φ of $\mathcal{L}_{\mathsf{t},\rhd}$, it is possible to define a function $\ulcorner\urcorner$ such that $\ulcorner\varphi\urcorner$ is a closed term of $\mathcal{L}_{\mathsf{t},\rhd}$, and $\ulcorner\urcorner$ is representable in $\mathcal{L}_{\mathsf{t},\rhd}$.

Condition 1 fixes a selection of logical constants. The connectives $\vee, \supset, \equiv, \exists$ can be defined as usual. Condition 2 completes the list of logical constants of $\mathcal{L}_{\mathsf{t},\rhd}$ by adding Tr and $\rhd$, the truth predicate, and the symbol for the evidential conditional. Condition 3 is necessary in order to express the function that associates each syntactic object to exactly one element of the domain. When there is no danger of confusion, we will identify an expression with its code.

As customary, we assume that $\mathcal{L}_{\mathsf{t},\rhd}$ has at least one countable acceptable model $\mathcal{M}$ with domain M. We also assume that, for every $a \in M$, there is a constant c_a in $\mathcal{L}_{\mathsf{t},\rhd}$. These two assumptions are completely harmless, since virtually any language suitable for the development of theories of truth can easily satisfy them. A full definition of the acceptability requirement would be too long, but it suffices to say that this requirement ensures the smooth encoding of syntax that is needed to develop a theory of truth.[13] The

[12] This is the kind of definition provided in [4, 21, 32].

[13] We refer to [27], Chapter 5. The key idea is that $\mathcal{M}$ is acceptable when it enables us to define a *coding scheme* and a *decoding scheme*. The former consists in an isomorphic copy of the natural numbers with

requirement that we have a constant for each element in M simplifies the interpretation of the quantifiers in the semantic construction of section 4.

Finally, we assume that the requirement for the so-called *strong diagonalization* is satisfied, i.e. that for every formula φ where a variable x is free, there is a term t such that $\mathcal{M} \models t = \ulcorner(\varphi)^t_x\urcorner$, where $(\varphi)^t_x$ is the result of uniformly replacing every free occurrence of x with t in φ. This condition ensures the construction of diagonal formulae, thus modeling the sentences employed in semantic paradoxes such as the Liar, or similar self-referential sentences. More specifically, a Liar sentence is of the form $\neg\mathsf{Tr}t_\lambda$ where $\mathcal{M} \models t_\lambda = \ulcorner\neg\mathsf{Tr}t_\lambda\urcorner$. Similarly, a *truth-teller* sentence has the form $\mathsf{Tr}t_\tau$ where $\mathcal{M} \models t_\tau = \ulcorner\mathsf{Tr}t_\tau\urcorner$. Following the customary use, we will abbreviate $\neg\mathsf{Tr}t_\lambda$ as λ and $\mathsf{Tr}t_\tau$ as τ.

In the next two sections we will define a model for $\mathcal{L}_{\mathrm{t},\rhd}$ by means of a construction that integrates the evidential account of conditionals with a Kripkean treatment of truth predications. Before that, it may be useful to provide some methodological remarks.

From a semantic point of view, the two ingredients of our construction have quite different features. Kripkean truth is extensional in at least two respects. First, in their original formulation, the satisfaction conditions for a formula that contains the truth predicate are given in a model without relativization to further coordinates. Second, the truth predicate enjoys a form of substitutivity *salva veritate*: for any sentence φ, any sentence that results from φ by replacing a sub-sentence ψ with $\mathsf{Tr}\ulcorner\psi\urcorner$ or *vice versa* receives the same semantic value as φ. The evidential conditional, by contrast, is non-extensional in both respects. First, the truth conditions for $\varphi \rhd \psi$ are given in a model relative to worlds in such a way that, for each world w, the value of $\varphi \rhd \psi$ in w depends on the values of φ and ψ in other worlds. Second, even if χ and φ have the same value in w, this does not guarantee that the same goes for $\varphi \rhd \psi$ and $\chi \rhd \psi$, unless χ and φ are logically equivalent. Similarly, even if χ and ψ have the same value in w, this does not guarantee that the same goes for $\varphi \rhd \psi$ and $\varphi \rhd \chi$, unless, again, χ and ψ are logically equivalent.

How can these two notions be combined within a coherent and well-defined semantics? Our idea is to construct a model for $\mathcal{L}_{\mathrm{t},\rhd}$ in such a way that its set of worlds is akin to a set of Kripkean fixed points, so that a formula $\varphi \rhd \psi$ can be evaluated in the model relative to such worlds by applying the Chrysippus Test as illustrated in section 2. Of course, fixed points must be suitably defined, otherwise there is no guarantee that $\varphi \rhd \psi$ belongs to a world w when it is true in w according to Definition 2.1, or that $\neg(\varphi \rhd \psi)$ belongs to w when it is false in w according to Definition 2.2. Moreover, naïveté must be validated, in the sense that the value of $\varphi \rhd \psi$ in w must correspond to the value of $\mathsf{Tr}\ulcorner\varphi \rhd \psi\urcorner$ in w.

In order to combine Kripke's truth-theoretic construction with a semantics for a conditional, there are at least two main routes, which correspond to well-known

their associated ordering, and an injective mapping from finite sequences (of any length, including 0) of elements of M to M itself. The latter consists in the set of finite sequences, the function that associates with each finite sequence its length, and the projection function that associates with each sequence s and each natural number n, the element occupying the n-th place in s. Since the elements of the syntax (alphabet, terms, formulae) are strings of symbols, if we associate each component of such strings to an element of M, by the acceptability requirement we can also associate strings of such elements (and hence complex syntactic objects such as terms and formulae) to elements of M.

approaches in the literature. The first is the revision-theoretic route. One starts by building Kripkean fixed points for $\mathcal{L}_{t,\rhd}$ where conditionals are interpreted arbitrarily. Then, one implements a modal valuation clause which revises the value of the conditional. Now conditionals have a new value, but truth predications involving conditionals and their compounds cannot be suitably interpreted. Therefore, one builds another Kripkean fixed point, starting from the revised values of the conditionals, and so on, until the revision procedure converges on stable values or (as is more often the case) the revision values start cycling. This idea has been systematically pursued by Yablo and Field in order to integrate logically strong conditionals and naïve truth.[14] The second route calls for a modification of Kripke's construction where the valuation clauses for the modal notions (in our case Definitions 2.1 and 2.2) are internalized. This approach has been adopted by Halbach, Leitgeb, Welch, and Stern in order to provide Kripke-style fixed-point semantics for necessity and possibility predicates.[15] Here we will take the second route, leaving to future investigation the question of whether the first one is feasible.

4 The Evidential Kripke Construction

Let $\mathcal{L}$ be the fragment of $\mathcal{L}_{t,\rhd}$ that contains neither $\rhd$ nor Tr. A *literal* is an atomic or negated atomic sentence of $\mathcal{L}$. Our construction starts from a set of *base worlds* defined as follows:[16]

Definition 4.1 W is a countable set of *base worlds*, that is, maximal consistent sets of literals.

In order to represent the relation of comparative distance that is required for the truth conditions of the evidential conditional, our base worlds are assumed to be ordered.

Definition 4.2 For every $w \in W$, let $\preceq_w$ be a binary relation that satisfies the following conditions:

(i) for every $w', w'', w''' \in W$, if $w' \preceq_w w''$ and $w'' \preceq_w w'''$, then $w' \preceq_w w'''$;
(ii) for every $w', w'' \in W$, either $w' \preceq_w w''$ or $w'' \preceq_w w'$;
(iii) for every $w' \in W$, $w \preceq_w w'$.

Let $S(\preceq)$ be the set of all such relations.

Informally speaking, $w' \preceq_w w''$ means that, from the point of view of w, w' is at least as close as w''. Accordingly, its negation, $w' \npreceq_w w''$ means that, from the point of view of w, w'' is strictly closer than w'. (i) says that $\preceq_w$ is transitive. (ii) says that $\preceq_w$ is strongly connected. (iii) says that $\preceq_w$ includes w at its minimum, although it may not be the only world in that position. We will call *w-minimal* any w' such that $w' \preceq_w w''$ for every w''.[17]

[14] Yablo [40], Field [9–14].

[15] Halbach et al. [19, 20], and Stern [37, 38].

[16] The requirement that W is countable can be relaxed, but we keep it for simplicity.

[17] Crupi and Iacona [4] employs centered systems of spheres, where centering is stronger than w-minimality as required by (iii). But (i)-(iii) will suffice for our purposes.

Now we have a set of base worlds W and a binary relation on them. But base worlds are just sets of literals, which do not include complex formulae nor truth predications. As anticipated above, we need to provide a model-theoretic construction that gradually completes, as it were, each of the base worlds with complex formulae and truth predications. More precisely, given the set W, our construction will build a set W^+ such that, for each $w_i \in W$, W^+ includes a set w_i^+ that includes w_i, and also contains complex formulae and truth predications based on the literals in W. The set w_i^+ is a *world stage*, i.e. the first stage based on w_i. This process of progressively adding world stages — $W_\alpha, W_{\alpha+1}, \ldots$ — eventually leads to *full worlds*, that is, worlds that include all sentences that can be included, given the base worlds and the valuation clauses for the vocabulary. Full worlds, therefore, behave as fixed points: iterating the process of adding world stages to full worlds results in the very same full worlds. Full worlds include base worlds as subsets, but also contain (codes of) sentences of the full language $\mathcal{L}_{\mathsf{t},\rhd}$.

How do we go from a world stage to the next? For sentences whose main logical operator is $\neg, \wedge, \forall$, as well as truth predications, determining which sentences are in w_i^+ only requires one to move down vertically, as it were, and look at the formulae in w_i. By contrast, in order to determine whether $\varphi \rhd \psi$ is in w_i^+, one needs to move not only vertically (i.e. seeing whether φ and/or ψ (and/or their negation) are in w_i), but horizontally as well, i.e. seeing what happens to φ and ψ in the other worlds. This figurative talk of 'vertical' and 'horizontal' movement can be made precise: vertical shifts concern different ordinal stages of the same world, while horizontal shifts concern different worlds at the same ordinal stage.

Note that the horizontal movement will also employ information about the comparative distance holding not just amongst base worlds, but also amongst world stages. We assume that the ordering on base worlds is preserved across all world stages. The rationale for this assumption is that the comparative distance between worlds is uniquely determined by non-semantic, atomic facts. In particular, this means that the properties of such ordering, specified in Definition 4.2 — and notably, w-minimality — are preserved across every stage of the construction.

From a technical point of view, the construction requires that the step from W_α to $W_{\alpha+1}$ is ruled by a *positive elementary definition*, which yields a *monotonic jump*, thus ensuring the required fixed-point properties of full worlds.[18] Since we need to refer to all the elements of the set of world stages W_α in order to define $W_{\alpha+1}$, it is convenient to define our jump on functions that assign world stages to natural numbers.[19] More specifically, consider the function $\mathsf{F_b} : \omega \longmapsto \mathcal{P}(M)$ such that $\mathsf{F_b}[\omega] = W$. In other words, for every $i \in \omega$, $\mathsf{F_b}(i) = w_i$. The intended jump from W to W^+ is then modelled as a jump on such a function $\mathsf{F_b}$, i.e. as a function that yields a function $\mathsf{F_b^+}$ such that (1) for every $i \in \omega$, $\mathsf{F_b}(i) \subseteq \mathsf{F_b^+}(i)$, i.e. for every index i, $\mathsf{F_b^+}$ yields world stages that include the base worlds yielded by $\mathsf{F_b}$, and (2) $\mathsf{F_b^+}(i)$ also includes complex sentences and truth predications, which follow the strong Kleene schema for $\neg$, $\wedge$,

[18] For the general theory of inductive definitions, see [27]. For their application to Kripkean theories of truth, see e.g. [26], Chapter 5.

[19] This is because W is assumed to be countable, and so is each W_α and the set of full worlds. If W is taken to be uncountable, the domain of the function is an appropriately sized index set.

$\forall$, the naïveté intuition for Tr, and Definitions 2.1 and 2.2 for $\rhd$. The process is then iterated until one reaches a fixed point $\mathsf{F}_\mathsf{b}^\delta$ such that, for every $i \in \omega$, $\mathsf{F}_\mathsf{b}^\delta(i)$ is a full world. For the sake of generality, however, the jump is not defined only for the specific function F_b that yields the base worlds, but on an arbitrary function $F : \omega \longmapsto \mathcal{P}(M)$. The resulting set of fixed points makes it then easily possible to specify $\mathsf{F}_\mathsf{b}^\delta$ and other functions delivering extensionally different full worlds.

Definition 4.3 Let F be a function from ω to $\mathcal{P}(M)$. The function F^+ is defined so that, for every $i \in \omega$ and $\varphi \in \mathcal{L}_{\mathsf{t},\rhd}$, $\varphi \in F^+(i)$ if:

1. $\varphi \in F(i)$, or
2. φ is a literal ψ and $\psi \in w_i$, or
3. φ is $\neg\neg\psi$ and $\psi \in F(i)$, or
3. φ is $\psi \wedge \chi$ and $\psi \in F(i)$ and $\chi \in F(i)$, or
4. φ is $\neg(\psi \wedge \chi)$ and $\neg\psi \in F(i)$ or $\neg\chi \in F(i)$, or
5. φ is $\forall x\psi$ and for every closed term t, $(\psi)^t_x \in F(i)$, or
6. φ is $\neg\forall x\psi$ and for some closed term t, $\neg(\psi)^t_x \in F(i)$, or
7. φ is $\mathsf{Tr}\ulcorner\psi\urcorner$ and $\psi \in F(i)$, or
8. φ is $\neg\mathsf{Tr}\ulcorner\psi\urcorner$ and $\neg\psi \in F(i)$, or
9. φ is $\psi \rhd \chi$ and either for every $j \in \omega$, $\neg(\psi \wedge \neg\chi) \in F(j)$ and the following conditions hold:

   (a) for some $j \in \omega$, $\psi \in F(j)$;
   (b) for some $j \in \omega$, $\neg\chi \in F(j)$;

   or some $j \in \omega$ is such that $\psi \wedge \neg\chi \in F(j)$ and the following conditions hold:

   (c) for some $j \in \omega$, $F(j)$ is $F(i)$-minimal and either $\psi, \chi \in F(j)$ or $\neg\psi, \neg\chi \in F(j)$;
   (d) for every $j \in \omega$, either ($\neg\psi \in F(j)$ or $\chi \in F(j)$), or there is a $k \in \omega$ such that $F(j) \not\preceq_{F(i)} F(k)$ and $\psi, \chi \in F(k)$;
   (e) for every $j \in \omega$, either ($\neg\psi \in F(j)$ or $\chi \in F(j)$), or there is a $k \in \omega$ such that $F(j) \not\preceq_{F(i)} F(k)$ and $\neg\psi, \neg\chi \in F(k)$; or

10. φ is $\neg(\psi \rhd \chi)$ and either for every $j \in \omega$, $\neg(\psi \wedge \neg\chi) \in F(j)$ and at least one of the following conditions hold:

   ($-a$) for every $j \in \omega$, $\neg\psi \in F(j)$;
   ($-b$) for every $j \in \omega$, $\chi \in F(j)$;

   or there is a $j \in \omega$ such that $\psi \wedge \neg\chi \in F(j)$ and one of the following conditions holds:

   ($-c$) for every $F(i)$-minimal $F(j)$, either $\psi, \neg\chi \in F(j)$ or $\neg\psi, \chi \in F(j)$;
   ($-d$) for some $j \in \omega$, $\psi, \neg\chi \in F(j)$ and for every $k \in \omega$, either $F(j) \preceq_{F(i)} F(k)$, or $\neg\psi, \neg\chi \in F(k)$, or $\psi, \neg\chi \in F(k)$, or $\neg\psi, \chi \in F(k)$;
   ($-e$) for some $j \in \omega$, $\psi, \neg\chi \in F(j)$ and for every $k \in \omega$, either $F(j) \preceq_{F(i)} F(k)$, or $\psi, \chi \in F(k)$, or $\psi, \neg\chi \in F(k)$, or $\neg\psi, \chi \in F(k)$.

Clauses 1-8 are as in Kripke's standard construction: they define the jump for $\neg$, $\wedge$, $\forall$, and Tr.[20] Clauses 9 and 10 define the jump for $\rhd$. Clause 9 reproduces the truth conditions of $\psi \rhd \chi$ as given in Definition 2.1. The first disjunct expresses the idea that ψ and $\neg\chi$ are absolutely incompatible, that is, their joint truth is impossible in spite of the fact that ψ is not impossible and χ is not necessary. The second disjunct expresses the idea that ψ and $\neg\chi$ are relatively incompatible, in the sense specified by conditions (c)-(e). In particular, (d), in the context of the above definition, might be more naturally spelled out as follows: 'for every $j \in \omega$ such that $\psi, \neg\chi \in F(j)$, there is $k \in \omega$ such that ...', but this would make membership in F appear in the antecedent of a material conditional (in the meta-language in which Definition 4.3 is given), thereby violating the positive elementary nature of the definition. In order to avoid this, we have 'internalized' the negation. Similar considerations hold for (e).[21] Clause 10 reproduces in similar way the falsity conditions of $\psi \rhd \chi$ as given in Definition 2.2.

5 Full Worlds as Fixed Points

Now it has to be shown how a set of full worlds is obtained from W. Let us start with a fundamental property of Definition 4.3.

Lemma 5.1 *Definition 4.3 is positive elementary in the relations $\preceq_w \in S(\preceq)$, with parameters in W.*

Proof Clauses 1-8 are immediately seen to be positive elementary, as they mimic those of Kripke's original construction, with parameters in W. Clauses 9 and 10, which concern $\rhd$, are also positive elementary: membership in F never appears in the scope of an odd number of negation symbols. Moreover, Definition 4.3 is positive elementary in the relations $\preceq_w \in S(\preceq)$, which justifies their negative occurrence in 9 and 10. Finally, the w-minimality condition is also positive, as it is formalized as $\forall w'(w' \preceq_w w'')$. □

Definition 4.3 characterizes a function that associates an input function $F : \omega \longmapsto \mathcal{P}(M)$ with an output function $F^+ : \omega \longmapsto \mathcal{P}(M)$. Since we are interested in iterations of this function, for notational convenience we now associate an operator acting on tuples of subsets of M with Definition 4.3. Let $\zeta(n, F)$ abbreviate the right-hand side of Definition 4.3, and let $[\omega \rightarrow \mathcal{P}(M)]$ denote the function space between ω and $\mathcal{P}(M)$.

[20] Strictly speaking, they are a variant of Kripke's original construction, from [16], Definition 15.5, but the difference is inessential.

[21] Note first that the non-positive clause 'for every $j \in \omega$ such that $\psi, \neg\chi \in F(j)$, there is $k \in \omega$ such that ...' is equivalent to the (non-positive) 'for every $j \in \omega$ (either $\psi, \neg\chi \notin F(j)$, or there is $k \in \omega$ such that ...)'. The first disjunct above is a negated conjunction: '$\psi \notin F(j)$ and $\neg\chi \notin F(j)$'. It is therefore equivalent to 'not $\psi \notin F(j)$ or not $\neg\chi \notin F(j)$.' But, now, we can 'internalize' the meta-linguistic negation, and apply it to the formulae ψ and $\neg\chi$ respectively, obtaining our item (d). Of course, the 'internalized' clause is not equivalent to the negative clause, but it yields (unlike the latter) only positive occurrences of membership in F. It is, therefore, the 'positive' version of condition (d) of Definition 2.1, and no difference arises from its non-positive counterpart when only classical values are involved.

Definition 5.2 Let $\Phi : [\omega \to \mathcal{P}(M)] \longmapsto [\omega \to \mathcal{P}(M)]$ be the operator such that $\Phi(F) := \{n \in M \mid \zeta(n, F)\}$ (n codes φ in Definition 4.3).

Call Φ the *evidential Kripke jump*. We now state two basic properties of Φ, namely, that it is *inclusive* and *monotone*:

Lemma 5.3 *For every $F, F' \in [\omega \to \mathcal{P}(M)]$:*

(i) $F \leq \Phi(F)$;
(ii) *If* $F \leq F'$, *then* $\Phi(F) \leq \Phi(F')$

where $F \leq F'$ means that, for every $i \in \omega$, $F(i) \subseteq F'(i)$.

Proof Disjunct 1 of Definition 4.3 guarantees (i). (ii) follows from Lemma 5.1, as positive elementary definitions yield monotone operators.[22] □

The monotonicity of Φ ensures that it has *fixed points*, i.e. that there are functions $F \in [\omega \to \mathcal{P}(M)]$ such that $\Phi(F) = F$, and that such fixed points form a lattice (Knaster-Tarski Theorem). Therefore, in particular, there is a *least* fixed point of Φ, call it F_Φ, such that for every fixed point F, $\mathsf{F}_\Phi \leq F$.

The least fixed point I_Φ can be more informatively described via iterations of Φ along the ordinals, as per the next definition.

Definition 5.4 For every ordinal α and every $F \in [\omega \to \mathcal{P}(M)]$, let the α-th iteration of Φ on F be defined by transfinite induction as follows.[23] For every successor ordinal $\alpha + 1$:

$$\Phi^{\alpha+1}(F) = \Phi(\Phi^{\alpha}(F))$$

For every limit ordinal δ:

$$\Phi^{\delta}(F) = \bigcup_{\beta<\delta} \Phi^{\beta}(F)$$

It is also convenient to isolate the stages that result from iterated applications of Φ. For every ordinal α, put $F^{\alpha} := \Phi^{\alpha}(F)$. We say that F^{α} is the α-th stage of F.

Finally, we can define the result of all possible iterations of Φ, i.e. the result of iterating Φ along all the ordinals:

$$\mathsf{F}_\Phi(F) := \bigcup_{\alpha \in \mathsf{Ord}} \Phi^{\alpha}(F).$$

Let $\mathsf{F}_\varnothing$ be the function from ω to $\mathcal{P}(M)$ such that, for every $i \in \omega$, $\mathsf{F}_\varnothing(i) = \varnothing$. It is easy to see that

$$\mathsf{F}_\Phi = \bigcup_{\alpha \in \mathsf{Ord}} \Phi^{\alpha}(\mathsf{F}_\varnothing),$$

thus obtaining a description of the least fixed point via all the possible iterations of Φ on the function $\mathsf{F}_\varnothing$ that treats every possible world as empty.

This characterization of I_Φ is useful to prove the next result. Let's say that a fixed point F is consistent if there is no $i \in \omega$ and $\varphi \in \mathcal{L}_{\mathsf{t},\triangleright}$ such that $\varphi \in F(i)$ and $\neg\varphi \in F(i)$.

[22] See [27], Chapter 1.

[23] We employ a class-formulation of transfinite induction; see, e.g., [22], Chapter 2.

Proposition 5.5 F_Φ *is consistent.*

Proof Let $\mathsf{F}^\alpha_\Phi = \Phi^\alpha(\mathsf{F}_\varnothing)$ We show by transfinite induction that there is no least ordinal α such that F^α_Φ is inconsistent, which entails that F_Φ is consistent.

Basis: $\alpha = 0$ or $\alpha = 1$. In the first case, for every $i \in \omega$, $\mathsf{F}^0_\Phi(i) = \mathsf{F}_\varnothing(i) = \varnothing$. In the second, by clause 2 of Definition 4.3, for every $i \in \omega$, $\mathsf{F}^1_\Phi(i) = w_i$. By Definition 4.1, each w_i is a maximally consistent set of literals. So in both cases the claim is true.

Successor case: $\alpha = \alpha_0 + 1$ ($\alpha_0 \geq 1$) and assume as inductive hypothesis that the claim holds up to α_0. Suppose that α is the least ordinal such that F^α_Φ is inconsistent. Then, there is an $i \in \omega$ and a $\varphi \in \mathcal{L}_{\mathrm{t},\rhd}$ such that $\varphi \in \mathsf{F}^\alpha_\Phi(i)$ and $\neg\varphi \in \mathsf{F}^\alpha_\Phi(i)$. Moreover, $\mathsf{F}^\alpha_\Phi = \Phi(\mathsf{F}^{\alpha_0}_\Phi)$ and by α's minimality, $\mathsf{F}^{\alpha_0}_\Phi$ is consistent. We now show how this supposition leads to absurdity.

We only consider the case in which φ has the form $\psi \rhd \chi$ (the other cases are routine). Suppose that $\psi \rhd \chi \in \mathsf{F}^\alpha_\Phi(i)$ and $\neg(\psi \rhd \chi) \in \mathsf{F}^\alpha_\Phi(i)$. By α's minimality, $\psi \rhd \chi \notin \mathsf{F}^{\alpha_0}_\Phi(i)$ or $\neg(\psi \rhd \chi) \notin \mathsf{F}^{\alpha_0}_\Phi(i)$. Suppose (without loss of generality) that $\psi \rhd \chi \notin \mathsf{F}^{\alpha_0}_\Phi(i)$. Then, by α's minimality, α is the least ordinal such that $\psi \rhd \chi \in \mathsf{F}^\alpha_\Phi(i)$, and hence clause 9 of Definition 4.3 applies to $\mathsf{F}^{\alpha_0}_\Phi(i)$. Clause 9 is a disjunction, so there are two cases.

Case 1: the first disjunct holds. In this case, for every $j \in \omega$, $\neg(\psi \wedge \neg\chi) \in \mathsf{F}^{\alpha_0}_\Phi(j)$, and (a) and (b) hold, i.e. there is a $k \in \omega$ such that $\psi \in \mathsf{F}^{\alpha_0}_\Phi(k)$ and there is an $l \in \omega$ such that $\neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(l)$. Since $\neg(\psi \rhd \chi) \in \mathsf{F}^{\alpha_0}_\Phi(i)$, there is an ordinal $\delta + 1 \leq \alpha_0$ such that $\neg(\psi \rhd \chi) \notin \mathsf{F}^\delta_\Phi(i)$ but $\neg(\psi \rhd \chi) \in \mathsf{F}^{\delta+1}_\Phi(i)$. So, two cases must be considered, which correspond respectively to the first and to the second disjunct of clause 10 of Definition 4.3. If the first disjunct holds, for every $j \in \omega$, $\neg(\psi \wedge \neg\chi) \in \mathsf{F}^\delta_\Phi(j)$ and (-a) or (-b) applies, that is, for every $k \in \omega$, $\neg\psi \in \mathsf{F}^\delta_\Phi(k)$ or for every $l \in \omega$, $\chi \in \mathsf{F}^\delta_\Phi(l)$. In this case, since Φ is inclusive (Lemma 5.3), we get that either $\neg\psi \in \mathsf{F}^{\alpha_0}_\Phi(k)$, which contradicts α's minimality by (a), or that $\chi \in \mathsf{F}^{\alpha_0}_\Phi(l)$, which contradicts α's minimality by (b). If the second disjunct holds, there is a $k \in \omega$ such that $\psi \wedge \neg\chi \in \mathsf{F}^\delta_\Phi(k)$. Since Φ is inclusive, $\mathsf{F}^\delta_\Phi(k) \subseteq \mathsf{F}^{\delta+1}_\Phi(k) \subseteq \mathsf{F}^{\alpha_0}_\Phi(k)$, and hence $\psi \wedge \neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(k)$. By our initial assumption, $\neg(\psi \wedge \neg\chi) \in \mathsf{F}^{\alpha_0}_\Phi(k)$, thus contradicting again α's minimality.

Case 2: for some $j \in \omega$, $\psi \wedge \neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(j)$ and clauses (c)-(e) hold. In this case, the following holds:

(1) by (c), for some $j \in \omega$, $\mathsf{F}^{\alpha_0}_\Phi(j)$ is $\mathsf{F}^{\alpha_0}_\Phi(i)$-minimal and $\psi, \chi \in \mathsf{F}^{\alpha_0}_\Phi(j)$ or $\neg\psi, \neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(j)$,

(2) by (d), for every $j \in \omega$, either $\neg\psi \in \mathsf{F}^{\alpha_0}_\Phi(j)$ or $\chi \in \mathsf{F}^{\alpha_0}_\Phi(j)$, or there is a $k \in \omega$ such that $\mathsf{F}^{\alpha_0}_\Phi(j) \not\preceq_{\mathsf{F}^{\alpha_0}_\Phi(i)} \mathsf{F}^{\alpha_0}_\Phi(k)$ and $\psi, \chi \in \mathsf{F}^{\alpha_0}_\Phi(k)$,

(3) by (e), for every $j \in \omega$, either $\neg\psi \in \mathsf{F}^{\alpha_0}_\Phi(j)$ or $\chi \in \mathsf{F}^{\alpha_0}_\Phi(j)$, or there is a $k \in \omega$ such that $\mathsf{F}^{\alpha_0}_\Phi(j) \not\preceq_{\mathsf{F}^{\alpha_0}_\Phi(i)} \mathsf{F}^{\alpha_0}_\Phi(k)$ and $\neg\psi, \neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(k)$.

Since $\neg(\psi \rhd \chi) \in \mathsf{F}^{\alpha_0}_\Phi(i)$, there is an ordinal $\delta + 1 \leq \alpha_0$ such that $\neg(\psi \rhd \chi) \notin \mathsf{F}^\delta_\Phi(i)$ but $\neg(\psi \rhd \chi) \in \mathsf{F}^{\delta+1}_\Phi(i)$. Given that for some $j \in \omega$, $\psi \wedge \neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(j)$, and Φ is inclusive, the first disjunct of clause 10 of Definition 4.3 is excluded, and therefore the second one must hold, that is, there is a $k \in \omega$ such that $\psi \wedge \neg\chi \in \mathsf{F}^\delta_\Phi(k)$ and one of (-c), (-d), and (-e) holds. We consider these cases in turn:

(4) For every $\mathsf{F}^\delta_\Phi(i)$-minimal $\mathsf{F}^\delta_\Phi(m)$, either $\psi, \neg\chi \in \mathsf{F}^\delta_\Phi(m)$ or $\neg\psi, \chi \in \mathsf{F}^\delta_\Phi(m)$. Since Φ is inclusive and $\delta < \alpha_0$, either $\psi, \neg\chi \in \mathsf{F}^{\alpha_0}_\Phi(m)$ or $\neg\psi, \chi \in \mathsf{F}^{\alpha_0}_\Phi(m)$.

But the ordering is constant throughout world stages and therefore, by (1) above, $\mathsf{F}_\Phi^{\alpha_0}(j)$ is $\mathsf{F}_\Phi^{\alpha_0}(i)$-minimal, and either ψ and $\neg\psi \in \mathsf{F}_\Phi^{\alpha_0}(i)$ or χ and $\neg\chi \in \mathsf{F}_\Phi^{\alpha_0}(i)$, against α's minimality.

(5) Both the following hold:

(5.1) for some $m \in \omega$, $\psi, \neg\chi \in \mathsf{F}_\Phi^{\delta}(m)$, and
(5.2) for every $n \in \omega$, $\neg\psi, \neg\chi \in \mathsf{F}_\Phi^{\delta}(n)$, or $\psi, \neg\chi \in \mathsf{F}_\Phi^{\delta}(n)$, or $\neg\psi, \chi \in \mathsf{F}_\Phi^{\delta}(n)$, or $\mathsf{F}_\Phi^{\delta}(m) \preceq_{\mathsf{F}_\Phi^{\delta}(i)} \mathsf{F}_\Phi^{\delta}(n)$.

Again, since Φ is inclusive, $\delta < \alpha_0$, and the ordering is constant, we can 'lift' (5.1) and (5.2) to α_0, i.e.:

(5.3) $\psi, \neg\chi \in \mathsf{F}_\Phi^{\alpha_0}(m)$, and
(5.4) for every $n \in \omega$, $\neg\psi, \neg\chi \in \mathsf{F}_\Phi^{\alpha_0}(n)$, or $\psi, \neg\chi \in \mathsf{F}_\Phi^{\alpha_0}(n)$, or $\neg\psi, \chi \in \mathsf{F}_\Phi^{\alpha_0}(n)$, or $\mathsf{F}_\Phi^{\alpha_0}(m) \preceq_{\mathsf{F}_\Phi^{\alpha_0}(j)} \mathsf{F}_\Phi^{\alpha_0}(n)$.

However, by (2):

(2.1) for every $m \in \omega$, either $\neg\psi \in \mathsf{F}_\Phi^{\alpha_0}(m)$ or $\chi \in \mathsf{F}_\Phi^{\alpha_0}(m)$, or
(2.2) for some $n \in \omega$, $\mathsf{F}_\Phi^{\alpha_0}(m) \not\preceq_{\mathsf{F}_\Phi^{\alpha_0}(i)} \mathsf{F}_\Phi^{\alpha_0}(n)$ and $\psi, \chi \in \mathsf{F}_\Phi^{\alpha_0}(n)$.

Both (5.3)-(2.1) and (5.4)-(2.2) contradict α's minimality.

(6) A contradiction obtains in a similar way (with (e)).

Limit case. This is immediate from the successor case. □

Paradoxical sentences such as the Liar yield a partial test for the inconsistency of fixed points that are obtained via iterations that take unions at limit stages:

Proposition 5.6 *For every $F \in [\omega \to \mathcal{P}(M)]$, if for some $j \in \omega$ and some ordinal α, either λ or $\neg\lambda \in F^\alpha(j)$, then $\mathsf{F}_\Phi(F)$ is inconsistent.*

Proof Suppose that $\lambda \in F^\alpha(j)$. Then, by clause 7 of Definition 4.3, $\mathsf{Tr}\ulcorner\lambda\urcorner \in F^{\alpha+1}(j)$. But since λ is equivalent to $\neg\mathsf{Tr}\ulcorner\lambda\urcorner$, by clause 1 we also get that $\neg\mathsf{Tr}\ulcorner\lambda\urcorner \in F^{\alpha+1}(j)$, and $F^{\alpha+1}(j)$ inconsistent. The case for $\neg\lambda$ is similar. □

Corollary 5.7 *For every $j \in \omega$, neither λ nor $\neg\lambda$ is in $\mathsf{F}_\Phi(j)$.*

Proof Directly from Propositions 5.5 and 5.6. □

Since F_Φ yields no inconsistent sets of sentences, we can take this function to provide *full worlds*.

Definition 5.8 The set W^Φ of *full worlds* is provided by all the values of F_Φ, i.e.

$$\mathsf{W}^\Phi := \mathsf{F}_\Phi[\omega]$$

Since Φ is inclusive, iterations of F_Φ extend each base world w to exactly one full world w^Φ. For every $i \in \omega$:

$$w_i = \mathsf{F}_\Phi^1(i) \subseteq \ldots \subseteq \mathsf{F}_\Phi(i) = \mathsf{w}_i^\Phi.$$

This makes precise our earlier informal claim that full worlds are "completions" of base worlds. Full worlds are the worlds we need in our combined theory of naïve truth and the evidential conditional. Since from now on we will only consider full worlds, in order to avoid notational cluttering, we will indicate them simply as w.

We conclude this section with a remark on the classification of "problematic" sentences. Kripke employed fixed points to define different kinds of such sentences. In particular, he called φ *ungrounded* when neither φ nor $\neg\varphi$ is in the least fixed point, and he called ψ *paradoxical* when neither ψ nor $\neg\psi$ is in any consistent fixed point. Since the least fixed point is consistent, it follows that every paradoxical sentence is ungrounded. For example, λ is both paradoxical and ungrounded. However, the converse does not hold: τ is ungrounded but not paradoxical, as there are non-minimal, consistent fixed points which include either τ or $\neg\tau$.

It is easy to see that our framework inherits the Kripkean taxonomy, as there are paradoxical and ungrounded sentences of $\mathcal{L}_{t,\rhd}$ which can be isolated by looking at non-minimal fixed points of the evidential Kripke construction, i.e. fixed points which strictly include full worlds. A simple example is the fixed point $\mathsf{F}_\Phi(F_\tau)$, where F_τ is the function from ω to $\mathcal{P}(M)$ such that, for every $i \in \omega$, $F^0(i) = \{\tau\}$. $\mathsf{F}_\Phi(F_\tau)$ can be shown to be consistent, and to include both τ and $\tau \rhd \tau$ in the sets it generates.

6 Evidential Kripke Frames and Valuations

Having shown how a set of full worlds can be obtained from our initial set of base worlds by means of an inductive definition, we are finally in a position to give a proper formulation of the semantics of $\mathcal{L}_{t,\rhd}$. Let us start with frames:

Definition 6.1 A *frame* for $\mathcal{L}_{t,\rhd}$ is a triple $\langle \mathcal{M}, \mathsf{W}^\Phi, S(\preceq)\rangle$, where $\mathcal{M}$ is a countable acceptable model of $\mathcal{L}$, W^Φ is the set of full worlds, and $S(\preceq)$ is the set of comparative distance relations between full worlds.

With each frame $\mathcal{F}$, one can associate a valuation function that assigns semantic values to sentences relative to full worlds.

Definition 6.2 For any frame $\mathcal{F}$, let v be a function from ordered pairs of elements of W^Φ and sentences of $\mathcal{L}_{t,\rhd}$ to the set $\{1, 1/2, 0\}$ such that, for every $w \in \mathsf{W}^\Phi$ and every $\varphi \in \mathcal{L}_{t,\rhd}$,

$$\mathsf{v}(w, \varphi) = \begin{cases} 1, & \text{if } \varphi \in w \\ 0, & \text{if } \neg\varphi \in w \\ 1/2, & \text{otherwise} \end{cases}$$

We call v the *valuation induced by* $\mathcal{F}$. Now we will state three important facts about valuations so defined.

Proposition 6.3 *For every frame $\mathcal{F}$, the valuation v induced by $\mathcal{F}$ is a* strong Kleene valuation. *That is, for every $w \in \mathsf{W}^\Phi$, $\varphi, \psi \in \mathcal{L}_{t,\rhd}$, and closed term t,*

(i) $\mathsf{v}(w, \neg\varphi) = 1 - \mathsf{v}(w, \varphi)$

(ii) $\mathsf{v}(w, \varphi \wedge \psi) = \min[\mathsf{v}(w, \varphi), \mathsf{v}(w, \varphi)]$
(iii) $\mathsf{v}(w, \forall x \psi) = \min[\mathsf{v}(w, (\psi)^t_x)]$

Proof The claim is immediate by inspection (the clauses of Definition 4.3 re-write the valuation clauses of strong Kleene semantics). □

Proposition 6.4 *For every frame $\mathcal{F}$, the induced valuation* v *satisfies the* intersubstitutivity of truth, *i.e. for every $w \in \mathsf{W}^{\Phi}$ and $\varphi \in \mathcal{L}_{\mathsf{t},\rhd}$, if φ^{t} is a formula resulting from φ by substituting (possibly non-uniformly) a sub-sentence ψ of φ with $\mathsf{Tr}\ulcorner\psi\urcorner$, or* vice versa, *then* $\mathsf{v}(w, \varphi) = \mathsf{v}(w, \varphi^{\mathsf{t}})$.

Proof First, note that for every frame $\mathcal{F}$, the induced valuation v satisfies *naïveté*, i.e. for every $w \in \mathsf{W}^{\Phi}$ and every $\varphi \in \mathcal{L}_{\mathsf{t},\rhd}$, $\mathsf{v}(w, \varphi) = \mathsf{v}(w, \mathsf{Tr}\ulcorner\varphi\urcorner)$. This is immediate from clause 7 of Definition 4.3 and Definition 5.8, by the fixed-point property of Φ. The result now follows by an easy induction, by the above claim and the compositionality of the clauses of Definition 4.3. □

As Propositions 6.3-6.4 show, our valuations interpret the standard logical vocabulary $\neg, \wedge, \forall$ as in strong Kleene semantics and treat the truth predicate in accordance with naïveté and intersubstitutivity of truth. Now it remains to be shown that they respect the truth and falsity conditions of the evidential conditional:

Proposition 6.5 *For every frame $\mathcal{F}$, the induced valuation* v *is such that, for every $w \in \mathsf{W}^{\Phi}$ and every $\varphi, \psi \in \mathcal{L}_{\mathsf{t},\rhd}$,*

- $\mathsf{v}(w, \varphi \rhd \psi) = 1$ *iff either every w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 0$ and the following conditions hold:*

  (a) $\mathsf{v}(w', \varphi) = 1$ *for some w';*
  (b) $\mathsf{v}(w', \psi) = 0$ *for some w';*

  or some w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 1$ and the following conditions hold:

  (c) either $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi) = 1$ *or* $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi) = 0$ *for some w-minimal w';*
  (d) for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg\psi) = 1$, *some strictly closer w'' is such that* $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi) = 1$;
  (e) for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg\psi) = 1$, *some strictly closer w'' is such that* $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi) = 0$.

- $\mathsf{v}(w, \varphi \rhd \psi) = 0$, *iff either every w' is such that* $\mathsf{v}(w', \varphi \wedge \neg\psi) = 0$ *and at least one of the following conditions hold:*

  (-a) $\mathsf{v}(w', \varphi) = 0$ *for every w';*
  (-b) $\mathsf{v}(w', \psi) = 1$ *for every w';*

  or there is w' such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 1$ *and at least one of the following conditions holds:*

  (-c) $\mathsf{v}(w', \varphi) = 1$ *and* $\mathsf{v}(w', \psi) = 0$, *or* $\mathsf{v}(w', \varphi) = 0$ *and* $\mathsf{v}(w', \psi) = 1$ *for every w-minimal w';*

(-d) *some* w' *is such that* $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg\psi) = 1$ *and no strictly closer* w'' *is such that* $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi) = 1$*;*

(-e) *some* w' *is such that* $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg\psi) = 1$ *and no strictly closer* w'' *is such that* $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi) = 0$.

- $\mathsf{v}(w, \varphi \rhd \psi) = 1/2$*, otherwise.*

Proof The claim directly follows from Definitions 4.3 and 6.2. The cases in which $\mathsf{v}(w, \varphi \rhd \psi) = 1$ are those in which $\varphi \rhd \psi \in w$ by clause 9 of Definition 4.3. The cases in which $\mathsf{v}(w, \varphi \rhd \psi) = 0$ are those in which $\neg(\varphi \rhd \psi) \in w$ by clause 10 of Definition 4.3. Finally, the case in which $\mathsf{v}(w, \varphi \rhd \psi) = 1/2$ is that in which $\varphi \rhd \psi \notin w$ and $\neg(\varphi \rhd \psi) \notin w$. □

Proposition 6.5 shows that, as far as the values 1 and 0 are concerned, the semantics of $\mathcal{L}_{t,\rhd}$ agrees with the truth and falsity conditions for the evidential conditional stated in section 1, namely, with Definitions 2.1 and 2.2. This result provides an extension of a well-known property of strong Kleene semantics, namely, that their valuation clauses are exactly as in classical logic — only, they are applied on the value space $\{1, 1/2, 0\}$ rather than $\{1, 0\}$ — and therefore when applied to formulae that have classical values yield the same result as the corresponding classical valuations.

A final remark concerns Identity. As explained in section 2, this principle fails for reasons that are independent of any consideration concerning paradoxical or ungrounded sentences, and can be expressed in the bivalent framework adopted by Crupi and Iacona. However, when the semantics is extended to the value space $\{1, 1/2, 0\}$ in order to deal with paradoxical and ungrounded sentences, there is a further argument to the effect that Identity should fail. The key assumption of the argument — call it *Non-Triviality* — is that any minimally interesting interpretation of $\rhd$ rules out the possibility that, for some sentences φ and ψ, the following four conditionals are all true: $\varphi \rhd \psi$, $\neg\varphi \rhd \psi$, $\varphi \rhd \neg\psi$, $\neg\varphi \rhd \neg\psi$. Having all of them true would entirely trivialize the notion of support, because it would imply that any member of the pair $\varphi, \neg\varphi$ is a reason for any member of the pair $\psi, \neg\psi$, thus depriving the word 'reason' of its intuitive meaning. In our semantic framework, Non-Triviality is at odds with Identity. As it turns out from Propostition 5.7, λ always gets value $1/2$, and the same goes for $\neg\lambda$. By clauses 9 and 10 of Definition 4.3 this entails that $\lambda \rhd \lambda$ also gets value $1/2$, and the same goes for $\neg\lambda \rhd \lambda$, $\lambda \rhd \neg\lambda$, $\neg\lambda \rhd \neg\lambda$. All these formulae are ungrounded (and, indeed, paradoxical). Now if Identity were valid, $\lambda \rhd \lambda$, $\neg\lambda \rhd \lambda$, $\lambda \rhd \neg\lambda$, $\neg\lambda \rhd \lambda$ would all be valid, against Non-Triviality. We regard this as a further reason for thinking that Identity should *not* hold unrestrictedly.

7 Logical Consequence

As is well known, strong Kleene valuations are compatible with different definitions of logical consequence, because there are different ways to extend the classical idea of truth preservation in every valuation to the value space $\{1, 1/2, 0\}$. In particular, a distinction can be drawn between being *strictly true*, which amounts to having value 1, and being *tolerantly true*, which amounts to having value 1 or 1/2.[24]

Definition 7.1 For every set Γ of $\mathcal{L}_{t,\rhd}$-sentences and frame $\mathcal{F} = \langle \mathcal{M}, \mathsf{W}^{\Phi}, S(\preceq) \rangle$, the valuation v induced by $\mathcal{F}$ makes Γ *strictly true* at w if it assigns 1 to every sentence in Γ at w, and it makes Γ *tolerantly true* at w if it assigns 1 or 1/2 to every sentence in Γ at w.

There are four direct ways to combine the notions of strict and tolerant truth into a definition of logical consequence, which yields four distinct logics: SS, TT, ST, TS. In the list below, these four combinations are applied to the semantics set out in the previous section, that is, the intended quantification is over sentences and sets of sentences of $\mathcal{L}_{t,\rhd}$, frames of the kind defined, and the valuations they induce.

Definition 7.2

$\Gamma \models_{ss} \varphi$ iff for any w, if Γ is strictly true in w, φ is strictly true in w;
$\Gamma \models_{tt} \varphi$ iff, for any w, if Γ is tolerantly true in w, φ is tolerantly true in w;
$\Gamma \models_{ts} \varphi$ iff, for any w, if Γ is tolerantly true in w, φ is strictly true in w;
$\Gamma \models_{st} \varphi$ iff, for any w, if Γ is strictly true in w, φ is tolerantly true in w.

Each of the four consequence relations so defined has its own distinctive features, and its theoretical advantages or disadvantages can be measured by different standards. However, we are not interested in comparing them. Here we will rely on SS, also known as *Strong Klenee Logic* or K3, to show that the logical properties of $\rhd$ stated in section 2 turn out to be preserved.[25] But similar results can be obtained by adopting TT, ST, or TS.

Proposition 7.3 $\varphi \rhd \psi \models_{ss} \varphi \supset \psi$ (*Material Implication* ✓)

Proof Assume that $\mathsf{v}(w, \varphi \rhd \psi) = 1$. By Proposition 6.5 this means that either every w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 0$ and conditions (a) and (b) are satisfied, or some w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 1$ and conditions (c)-(e) are satisfied. In the first case, $\mathsf{v}(w, \varphi \wedge \neg\psi) = 0$, so $\mathsf{v}(w, \varphi \supset \psi) = 1$. In the second case, given condition (iii) of Definition 4.2, it is not the case that $\mathsf{v}(w, \varphi) = 1$ and $\mathsf{v}(w, \psi) = 0$, which yields again that $\mathsf{v}(w, \varphi \supset \psi) = 1$. □

Proposition 7.4 $\varphi \rhd \psi, \varphi \rhd \chi \models_{ss} \varphi \rhd (\psi \wedge \chi)$ (*AND* ✓)

Proof Assume that $\mathsf{v}(w, \varphi \rhd \psi) = \mathsf{v}(w, \varphi \rhd \chi) = 1$. Since each of the two conditionals can be true in virtue of each of the disjuncts stated in Proposition 6.5, four cases are to be considered.

Case 1: both $\varphi \rhd \psi$ and $\varphi \rhd \chi$ are true in virtue of the first disjunct. In this case every w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = \mathsf{v}(w', \varphi \wedge \neg\chi) = 0$, some w' is such that $\mathsf{v}(w', \varphi) = 1$, some w' is such that $\mathsf{v}(w', \psi) = 0$, and some w' is such that $\mathsf{v}(w', \chi) = 0$. It follows

[24] We take this distinction and its uses from [2].

[25] This logic is commonly adopted as a starting point for the interpretation of the standard connectives, see [9, 11, 12, 24].

that every w' is such that $\mathsf{v}(w', \varphi \wedge \neg(\psi \wedge \chi)) = 0$, some w' is such that $\mathsf{v}(w', \varphi) = 1$, and some w' is such that $\mathsf{v}(w', \psi \wedge \chi) = 0$. Therefore, $\mathsf{v}(w, \varphi \rhd (\psi \wedge \chi)) = 1$.

Case 2: the first disjunct holds only for $\varphi \rhd \psi$. In this case (c)-(e) hold for $\varphi \rhd \chi$. (c) entails that, for some w-minimal w', either $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \chi) = 1$ or $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \chi) = 0$. If $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \chi) = 1$, then $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi \wedge \chi) = 1$, given that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 0$. If $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \chi) = 0$, then $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi \wedge \chi) = 0$. (d) entails that, for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(\neg\chi) = 1$, some strictly closer w'' is such that $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \chi) = 1$. Given the initial assumption about $\varphi \rhd \psi$, this yields that, for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg(\psi \wedge \chi)) = 1$, some strictly closer w'' is such that $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi \wedge \chi) = 1$. Moreover, from (e) we get that that, for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg\chi) = 1$, some strictly closer w'' is such that $\mathsf{v}(w'', \varphi) = \mathsf{v}(\chi) = 0$ which entails that, for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg(\psi \wedge \chi)) = 1$, some strictly closer w'' is such that $\mathsf{v}(w''\varphi) = \mathsf{v}(\psi \wedge \chi) = 0$. Therefore, $\mathsf{v}(w, \varphi \rhd (\psi \wedge \chi)) = 1$.

Case 3: the first disjunct holds only for $\varphi \rhd \chi$. This case is analogous to case 2.

Case 4: the first disjunct holds neither for $\varphi \rhd \psi$ nor for $\varphi \rhd \chi$. In this case (c)-(e) hold for both conditionals. From (c) we get that some w-minimal w' is such that either $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi) = 1$ or $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi) = 0$, and some w-minimal w'' is such that either $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \chi) = 1$ or $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \chi) = 0$. If either $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi) = 0$ or $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \chi) = 0$, then either $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi \wedge \chi) = 0$ or $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi \wedge \chi) = 0$. If $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi) = 1$ and $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \chi) = 1$, then $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \psi \wedge \chi) = 1$ and $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi \wedge \chi) = 1$ given (d). Moreover, from (d) we get that, for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg(\psi \wedge \chi)) = 1$, some strictly closer w'' is such that $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi \wedge \chi) = 1$. Finally, (e) yields that for every w' such that $\mathsf{v}(w', \varphi) = \mathsf{v}(w', \neg(\psi \wedge \chi)) = 1$, some strictly closer w'' is such that $\mathsf{v}(w'', \varphi) = \mathsf{v}(w'', \psi \wedge \chi) = 0$. □

Proposition 7.5 $\varphi \rhd \psi \models_{ss} \neg\psi \rhd \neg\varphi$ (*Contraposition* ✓)

Proof Assume that $\mathsf{v}(w, \varphi \rhd \psi) = 1$. By Proposition 6.5 this means that either every w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 0$ and conditions (a) and (b) are satisfied, or some w' is such that $\mathsf{v}(w', \varphi \wedge \neg\psi) = 1$ and conditions (c)-(e) are statisfied. In the first case, every w' is such that $\mathsf{v}(w', \neg\psi \wedge \neg\neg\varphi) = 0$, some w' is such that $\mathsf{v}(w', \neg\psi) = 1$, and some w' is such that $\mathsf{v}(w', \neg\varphi) = 0$. Therefore, $\mathsf{v}(w', \neg\psi \rhd \neg\phi) = 1$. In the second we get that (c)-(e) hold for $\neg\psi \rhd \neg\varphi$ as well. So, again, $\mathsf{v}(w', \neg\psi \rhd \neg\phi) = 1$. □

Proposition 7.6 $\varphi \rhd \chi \not\models_{ss} (\varphi \wedge \psi) \rhd \chi$ (*Monotonicity* ×)

Proof Let $W = \{w, w', w''\}$, $w' \not\preceq_w w$, $w'' \not\preceq_w w'$, and

$$\mathsf{v}(w, \varphi) = 1, \mathsf{v}(w, \psi) = 0, \mathsf{v}(w, \chi) = 1$$
$$\mathsf{v}(w', \varphi) = 0, \mathsf{v}(w', \psi) = 0, \mathsf{v}(w', \chi) = 0$$
$$\mathsf{v}(w''\varphi) = 1, \mathsf{v}(w'', \psi) = 1, \mathsf{v}(w''\chi) = 0$$

By Proposition 6.5, $\mathsf{v}(w, \varphi \rhd \chi) = 1$ because (c)-(e) are satisfied. Instead, $\mathsf{v}(w, (\varphi \wedge \psi) \rhd \chi) = 0$, given that neither (c) nor (d) are satisfied. □

Proposition 7.7 *Not:* $\varphi \rhd \psi \models_{ss} \varphi \rhd \chi$ *whenever* $\psi \models \chi$ (*Right Weakening* ×)

Proof Let $W = \{w, w', w''\}$, $w' \not\leq_w w$, $w'' \not\leq_w w'$, and

$$\mathsf{v}(w, \varphi) = 1, \mathsf{v}(w, \psi) = 1, \mathsf{v}(w, \chi) = 1$$
$$\mathsf{v}(w', \varphi) = 0, \mathsf{v}(w', \psi) = 1, \mathsf{v}(w', \chi) = 0$$
$$\mathsf{v}(w'', \varphi) = 1, \mathsf{v}(w'', \psi) = 0, \mathsf{v}(w'', \chi) = 1$$

By Proposition 6.5, $\mathsf{v}(w, \varphi \rhd (\psi \wedge \chi)) = 1$ because (c)-(e) are satisfied. However, $\mathsf{v}(w, \varphi \rhd \psi) = 0$ in spite of the fact that $\psi \wedge \chi \models \psi$. The reason is that (e) does not hold. □

Proposition 7.8 $\varphi \wedge \psi \not\models_{ss} \varphi \rhd \psi$ (*Conjunctive Sufficiency* $\times$)

Proof Let $W = \{w, w'\}$, $w' \not\leq_w w$, and

1. $\mathsf{v}(w, \varphi) = 1, \mathsf{v}(w, \psi) = 1$
2. $\mathsf{v}(w', \varphi) = 1, \mathsf{v}(w', \psi) = 0$

In this case $\mathsf{v}(w, \varphi \wedge \psi) = 1$. However, by Proposition 6.5 we get that $\mathsf{v}(w, \varphi \rhd \psi) = 0$ because (e) does not hold. □

Proposition 7.9 $\not\models_{ss} \varphi \rhd \varphi$ (*Identity* $\times$)

Proof Directly from Proposition 6.5, given that $\mathsf{v}(w, \varphi \rhd \varphi) = 0$ whenever $\mathsf{v}(w', \varphi) = 0$ for every w' or $\mathsf{v}(w', \varphi) = 1$ for every w'. □

Note that the proofs of Propositions 7.3-7.9 can easily be adapted to ST, given that strict truth suffices for tolerant truth. In order to provide analogous results in TT, instead, one should also reason under the hypothesis that the premises have the value $1/2$. Finally, although Propositions 7.3-7.9 are not provable in TS, where one cannot assume that the premises are strictly true in order to show that the conclusion is strictly true, in TS one can prove a meta-inferential version of these results. For example, in the case of Material Implication one can prove that if $\Gamma \models \varphi \rhd \psi$, then $\Gamma \models \varphi \supset \psi$, and similar formulations hold for the other facts about $\rhd$. So, as far as the key logical properties of $\rhd$ are concerned, there is no substantive difference between SS, TT, ST, and TS.

8 Validity and Paradox

Now we can provide a proper formulation of the Formalized Stoic Thesis. Let the two-place predicate Val be defined as follows, where $f_\rhd$ is a term for the recursive function such that, whenever x and y denote the (codes of) sentences φ and ψ, $f_\rhd$ returns the (code of the) sentence $\varphi \rhd \psi$:

Definition 8.1 $\mathsf{Val}xy \equiv_{\mathsf{def}} \mathsf{Tr} f_\rhd(x, y)$

Whenever x and y are replaced by closed terms $\ulcorner\varphi\urcorner$ and $\ulcorner\psi\urcorner$ (denoting the codes of φ and ψ, respectively), the formula $\mathsf{Val}\ulcorner\varphi\urcorner\ulcorner\psi\urcorner$ expresses the claim that the inference from φ to ψ is valid, which turns out to be equivalent to the claim that the conditional $\varphi \rhd \psi$ is true. In this section we illustrate some direct consequences of Definition 8.1 and discuss some of its implications.

Given Definition 8.1, validity inherits the logical properties of the evidential conditional. Consider Material Implication. From Proposition 7.3 we get that $\mathsf{Val}\ulcorner\varphi\urcorner\ulcorner\psi\urcorner$ entails $\varphi \supset \psi$. As is reasonable to expect, the claim that the inference from φ to ψ is valid is stronger than the mere negation of $\varphi \wedge \neg\psi$. Similar considerations hold for the other properties of $\rhd$ proved in section 7. From Propositions 7.4 and 7.5 we get that Val preserves AND and Contraposition. The case of Contraposition is particularly interesting here, since Crupi and Iacona regard Contraposition as a basic principle that holds independently of the distinction between conclusive and defeasible reasons. For example, consider the following arguments:

(5) Sophie can read French; therefore, she can read
(6) Sophie cannot read; therefore, she cannot read French
(7) Sophie is French; therefore, she can read French
(8) Sophie cannot read French; therefore, she is not French

While the reason stated by the premise of (5) is conclusive, in that it rules out the falsity of its conclusion, the reason stated by the premise of (7) is defeasible. But according to Crupi and Iacona an important analogy remains, in that in both cases it is plausible to expect that, if φ supports ψ, then $\neg\psi$ supports $\neg\varphi$. Thus, (6) and (8) seem no less compelling than (5) and (7) respectively.

Finally, from Propositions 7.6-7.9 we get that Val violates Monotonicity, Right Weakening, Conjunctive Sufficiency, and Identity. The failure of Monotonicity, in particular, shows that Val can be used to formally represent defeasible inferential relations.

A second point to be noted — and which is crucial to our project — is that ascriptions of validity turn out to be gappy in the same sense in which truth ascriptions are gappy. Since some truth predications involving conditionals receive the intermediate value $1/2$ due to semantic paradoxes, the same goes for the corresponding validity claims. In section 7 we saw that $\lambda \rhd \lambda$ always gets value $1/2$, so the same goes for $\mathsf{Tr}\ulcorner\lambda \rhd \lambda\urcorner$. Given Definition 8.1, this entails that $\mathsf{Val}\ulcorner\lambda\urcorner\ulcorner\lambda\urcorner$ also gets value $1/2$. In other words, as long as truth is understood as strict truth, one cannot truly claim that the following argument is valid:

(9) This sentence is not true; therefore, this sentence is not true.

Similarly, one cannot truly claim that (9) is *in*valid, for $\neg\mathsf{Val}\ulcorner\lambda\urcorner\ulcorner\lambda\urcorner$ has the same value as $\mathsf{Val}\ulcorner\lambda\urcorner\ulcorner\lambda\urcorner$. More generally, the theory of validity suggested here implies that, due to semantic paradoxes, some arguments are such that one can truly claim neither that they are valid nor that they are invalid. These arguments may be called *ungrounded*, just like the conditionals to which they correspond.

Note that, in our view, ungrounded arguments such as (9) differ from invalid arguments such as the following:

(10) $1 = 0$; therefore, snow is white.
(11) Snow is white; therefore, $0 = 0$.

The arguments (10) and (11) are invalid for the same reason for which the conditionals (1) and (2) are false, namely, that an impossible truth supports nothing, and nothing supports a necessary truth. The same goes for the following arguments:

(12) $1 = 0$; therefore, this sentence is not true.
(13) This sentence is true; therefore, $0 = 0$.

This shows that some arguments that contain ungrounded sentences are not themselves ungrounded: containing ungrounded sentences is not a sufficient condition for being ungrounded.

The class of ungrounded arguments include not only arguments in which one or more sentences are paradoxical (in our sense), but also claims about validity that are themselves paradoxical. Consider for example the following argument:

(14) The argument from this sentence to $0 \neq 0$ is valid; therefore, $0 \neq 0$.

This argument, which can be represented as a formula π equivalent to $\mathsf{Val}\ulcorner\pi\urcorner\ulcorner 0 \neq 0\urcorner$, has been analyzed by Beall and Murzi under the label *v-Curry*.[26] Beall and Murzi introduce a primitive, naïve notion of validity, formalized as a predicate and characterized by intuitive introduction and elimination rules (essentially, the introduction and elimination rules for the material conditional, only written for a predicate rather than a connective), and show that π gives rise to a version of Curry's Paradox that only employs structural rules, in addition to the validity rules.[27]

Although our validity predicate is not primitive, and we cannot obtain π by diagonalization from Definition 8.1, it is easy to see how π can be treated in our framework by means of some minimal adjustments. One option would be to set up a translation between our language $\mathcal{L}_{\mathsf{t},\triangleright}$ and a language $\mathcal{L}_{\mathsf{v}}$ defined as $\mathcal{L} \cup \{\mathsf{Val}\}$ using Kleene's Second Recursion Theorem, as explained by Halbach.[28] This way, π would be definable as the translation of a suitable $\mathcal{L}_{\mathsf{t},\triangleright}$-sentence, which would then turn out to be ungrounded. Alternatively, we could give our semantics directly for $\mathcal{L}_{\mathsf{v}}$, essentially formulating clauses 9 and 10 of Definition 4.3 for Val, rather than for $\triangleright$, and define Tr and $\triangleright$ via it. In fact, a similar strategy is pursued, in the context of naïve validity, by Nicolai and Rossi, but with the semantic clauses for Val matching the valuation clauses for the material conditional.[29] Regardless of which option we choose, π turns out to be neither true nor false, and (14) turns out to be neither valid nor invalid. It is therefore clear that our semantics extends the Kripkean idea of ungroundedness to arguments and their validity.

[26] Beall and Murzi [1]. Curry's Paradox is a conditional variant of the Liar which employs a sentence κ equivalent to $\mathsf{Tr}\ulcorner\kappa\urcorner \supset \bot$ (or $\mathsf{Tr}\ulcorner\kappa\urcorner \triangleright \bot$), see [35].

[27] Beall and Murzi argue that their naïve notion of validity is at least as well-motivated as naïve truth. As a consequence, the paradoxes of naïve validity are at least as urgent as the paradoxes of naïve truth, and the fact that no fully structural theory can block the v-Curry paradox (for the derivation of $0 \neq 0$ in it simply does not use any operational rule) puts pressure on the non-classical theorist to employ a substructural logic. For more on naïve validity, and the relation between naïve truth-theoretic and validity-theoretic principles and paradoxes, see [29].

[28] Halbach [16], §5.3. An accessible presentation of Kleene's Second Recursion Theorem and its uses can be found in [28].

[29] Nicolai and Rossi [30, 31].

9 Conclusions

In this paper we have shown how the idea that valid arguments amount to true conditionals can be accommodated within a consistent formal theory. In order to obtain the desired equivalence — the Formalized Stoic Thesis — we have combined naïve truth with the evidential conditional. As highlighted in section 1, the theory so obtained has at least two interesting implications.

First, the Formalized Stoic Thesis significantly increases our expressive resources, enabling us to formulate claims about validity that could not be formulated otherwise. These claims include ascriptions of validity to particular arguments, as in 'The argument from 'James is a bachelor' to 'James is unmarried' is valid', represented as $\mathsf{Val}\ulcorner Bt\urcorner, \ulcorner Ut\urcorner$, and generalizations, such as 'Every argument to the conclusion that $0 = 0$ is valid', i.e. $\forall x(\mathsf{Val}x\ulcorner 0 = 0\urcorner)$. Interestingly, the claims of the second kind can be used to express key properties of validity itself, such as 'Every argument of the form $\varphi \wedge (\varphi \supset \psi); \varphi$ is valid', which are usually stated in the meta-language. Certainly, there is a sense in which some such generalizations cannot be validated, since only their grounded instances are actually part of our theory. However, as we have observed, this form of partiality is a price worth paying in order to avoid triviality. So, the fact that our theory retains exactly the acceptable, non-problematic claims about validity, and express them in the object-language, seems more a virtue than a limitation. Note also that, in any Kripkean construction, not just in ours, gappy instances of desirable generalizations might not be validated.

Second, the Formalized Stoic Thesis yields a coherent unified treatment of the paradoxes of truth and validity, as we pointed out in section 8. Elaborating on our observations there, we notice that our theory already includes a treatment of naïve validity. As anticipated above, Nicolai and Rossi develop a Kripke-style semantics for a primitive validity predicate and interpret it essentially as a predicate version of the material conditional. As a consequence, the resulting naïve validity predicate obeys Beall and Murzi's introduction rules and avoids the v-Curry paradox by treating π (and all relevantly similar sentences) as ungrounded. Nicolai and Rossi call *grounded validity* the resulting theory of validity. It is easy to see that our theory recovers grounded validity because $\varphi \rhd \psi$ entails $\varphi \supset \psi$ (Proposition 7.3), and consequently $\mathsf{Tr}\ulcorner\varphi \rhd \psi\urcorner$ entails $\mathsf{Tr}\ulcorner\varphi \supset \psi\urcorner$. By Definition 8.1, $\mathsf{Tr}\ulcorner\varphi \rhd \psi\urcorner$ means that the argument from φ to ψ is valid. $\mathsf{Tr}\ulcorner\varphi \supset \psi\urcorner$, according to the valuation clauses of Nicolai and Rossi, just gives the extension of naïve validity (interpreted as grounded validity). Therefore, re-writing the former as a claim about validity in our sense and the latter as a claim about naïve validity, we get that validity in our sense entails naïve validity.

We close by pointing out two directions in which our work could be further developed. The first concerns the formal study of defeasible reasoning. The Formalized Stoic Thesis allows us to articulate a non-monotonic logic that models defeasible reasoning in the context of a first-order language with naïve truth. A distinctive principle of this logic is Contraposition, which we take to characterize validity broadly understood, as explained in Section 8. This is not a widespread assumption in the literature on non-monotonic logic, for Contraposition is usually *not* regarded as an essential property of defeasible reasoning.[30] But as far as we can see, no compelling reason has ever been provided for thinking that defeasible reasoning is non-contrapositive. In this respect, our proposal opens a perspective in which interesting logical results can be established.[31]

The second direction concerns the theory of truth. As it usually happens with Kripke-style constructions, several notions of logical consequence can be associated with the resulting models. We have focused on a restricted set of options, but there are still more to be considered. More specifically, supervaluationism and classical consequence (obtained via the so-called closing off of Kripkean fixed points) are cases in point. In both approaches, the logic would be significantly stronger, but one would lose the intersubstitutivity of truth. A supervaluational semantics could easily be associated with our account by adopting the elegant template developed by Stern.[32] A classical version of our theory, instead, would inherit much of the features of classical theories of Kripkean truth, such as Kripke-Feferman.[33] Such a version would be naturally compared with theories of object-linguistic validity formulated in classical logic.[34]

Acknowledgements We are grateful to Johannes Stern and to two anonymous referees for their valuable comments. We acknowledge the support of PLEXUS, a Marie Skłodowska-Curie action funded by the EU under the Horizon 2020 Research and Innovation Programme (Grant Agreement no 101086295).

Funding Open access funding provided by Università degli Studi di Torino within the CRUI-CARE Agreement.

References

1. Beall, J. C., & Murzi, J. (2013). Two flavors of Curry's paradox. *The Journal of Philosophy, 110*(3), 143–165.
2. Cobreros, P., Égré, P., Ripley, D., & van Rooij, R. (2012). Tolerant, classical, strict. *Journal of Philosophical Logic, 41*(2), 347–385.
3. Crupi, V., Dalmonte, T., & Iacona, A. (2022). Non-Monotonicity and Contraposition. *manuscript*.
4. Crupi, V., & Iacona, A. (2022). The Evidential Conditional. *Erkenntnis, 87*, 2897–2921.
5. Crupi, V., & Iacona, A. (2022). Three Ways of Being Non-Material. *Studia Logica, 110*, 47–93.
6. Crupi, V., & Iacona, A. (2024). Conditionals: Inferentialism Explicated. *Erkenntnis*, forthcoming.

[30] In the most influential works on non-monotonic logic, such as [15, 23], and others, Contraposition is never included among the basic principles.

[31] Crupi and Iacona [7] outline a general theory of reasons along these lines. Crupi et al. [3] develops the same line of thought in terms of a non-monotonic consequence relation.

[32] Stern [39].

[33] More specifically, one would lose inference rules to both introduce and eliminate the truth predicate and could consistently recover at most one such principle. See [8] and [16] (Ch. 15) for more details.

[34] Such as the one in [17, 18]

7. Crupi, V., & Iacona, A. (2023). Outline of a Theory of Reasons. *Philosophical Quarterly, 73*, 117–142.
8. Feferman, S. (1991). Reflecting on incompleteness. *The Journal of Symbolic Logic, 56*(1), 1–49.
9. Field, H. (2003). A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic, 32*(2), 139–177.
10. Field, H. (2007). Solving the paradoxes, escaping revenge. In Jc. Beall (Ed.), *Revenge of the Liar: new essays on the paradox* (pp. 78–144). Oxford: Oxford University Press.
11. Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.
12. Field, H. (2014). Naive truth and restricted quantification: Saving truth a whole lot better. *The Review of Symbolic Logic, 7*(1), 147–191.
13. Field, H. (2016). Indicative conditionals, restricted quantification, and naive truth. *The Review of Symbolic Logic, 9*(1), 181–208.
14. Field, H. (2020). Properties, propositions and conditionals. *Australasian Philosophical Review, 4*(2), 112–146.
15. Gabbay, D. M. (1985). Theoretical Foudation for Non-Monotonic Reasoning in Expert Systems. In: *Logic and Models of Concurrent Systems*. Springer.
16. Halbach, V. (2014). *Axiomatic theories of truth*. Second Edition. Cambridge University Press.
17. Halbach, V. (2020). Formal notes on the substitutional analysis of logical consequence. *Notre Dame Journal of Formal Logic, 61*(2), 317–339.
18. Halbach, V. (2020). The substitutional analysis of logical consequence. *Noûs, 54*(2), 431–450.
19. Halbach, V., Leitgeb, H., & Welch, P. (2003). Possible-worlds semantics for modal notions conceived as predicates. *Journal of Philosophical Logic, 32*(2), 179–223.
20. Halbach, V., & Welch, P. (2009). Necessities and necessary truths: A prolegomenon to the use of modal logic in the analysis of intensional notions. *Mind, 118*(469), 71–100.
21. Iacona, A. (2023). Valid Arguments as True Conditionals. *Mind, 132*, 428–451.
22. Jech, T. (2003). Set Theory. The Third Millennium Edition. Vol. 14. Berlin, Heidelberg, New York: Springer.
23. Kraus, S., Lehmann, D., & Magidor, M. (1990). Nonmonotonic reasoning, preferential models and cumulative logics. *Journal of Artificial Intelligence, 44*, 167–207.
24. Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy, 72*(19), 690–716.
25. Lewis, D. (1973). *Counterfactuals*. Blackwell.
26. McGee, V. (1991). *Truth, vagueness, and paradox: An essay on the logic of truth*. Hackett Publishing.
27. Moschovakis, Y. (1974). *Elementary Induction on Abstract Structures*. Amsterdam, London and New York: North-Holland / Elsevier.
28. Moschovakis, Y. (2010). Kleene's amazing second recursion theorem. *Bulletin of Symbolic Logic, 16*(2), 189–239.
29. Murzi, J., & Rossi, L. (2021). Naïve validity. *Synthese*, *199*(Suppl. 3), 819–841.
30. Nicolai, C., & Rossi, L. (2018). Principles for object-linguistic consequence: From logical to irreflexive. *Journal of Philosophical Logic, 47*(3), 549–577.
31. Nicolai, C., & Rossi, L. (2023). Systems for non-reflexive consequence. *Studia Logica, 111*, 947–977.
32. Raidl, E., Iacona, A., & Crupi, V. (2021). The Logic of the Evidential Conditional. *The Review of Symbolic Logic*, 1–13.
33. Rott, H. (1986). Ifs, though, and because. *Erkenntnis, 25*, 345–370.
34. Rott, H. (2019). Difference-making conditionals and the relevant Ramsey test. *Review of Symbolic Logic*.
35. Shapiro, L., & Beall, J. C. (2021). Curry's Paradox. In N. Edward (Ed.), *The Stanford Encyclopedia of Philosophy, Winter 2021.* Zalta. Metaphysics Research Lab: Stanford University.
36. Stalnaker, R. (1991). A Theory of Conditionals. In: Conditionals, edited by F. Jackson, 28–45. Oxford University Press.
37. Stern, J. (2014). Modality and axiomatic theories of truth I: Friedman-Sheard. *The Review of Symbolic Logic, 7*(2), 273–298.
38. Stern, J. (2014). Modality and axiomatic theories of truth II: Kripke-Feferman. *The Review of Symbolic Logic, 7*(2), 299–318.
39. Stern, J. (2018). Supervaluation-style truth without supervaluations. *Journal of Philosophical Logic, 47*(5), 817–850.

40. Yablo, S. (2004). New grounds for naive truth theory. In: Liars and Heaps. New Essays on Paradox, edited by Beall, Jc, 312–330. Oxford: Oxford University Press.