UNIVERSITAT DE BARCELONA

On Common Solutions to the Liar and the *Sorites*

Sergi Oms Sardans

On Common Solutions to the Liar and the *Sorites*

Sergi Oms Sardans

Phd in Cognitive Science and Language

Supervisor: José Martínez Fernández

Tutor: Manuel García-Carpintero Sánchez-Miguel

Department of Philosophy

Faculty of Philosophy

University of Barcelona

CONTENTS

ABSTRACT

In this dissertation I examine some of the most relevant proposals of common solutions to the Liar and the *Sorites* paradoxes. In order to do that, I present first a definition of what a paradox is so that, with this at hand, I can characterize in detail what should we expect from a common solution to a given collection of paradoxes. Next, I look into the reasons we might have to endorse a common solution to a group of paradoxes and some consequences are drawn with respect to Vann McGee's and Graham Priest's proposals to cope with both the Liar and the *Sorites* paradoxes. In the next chapters, three authors are examined in some detail. First, Jamie Tappenden's account is judged inappropriate, specially in the case of the Liar paradox. With respect to the *Sorites*, it is showed to be at least as problematic as Supervaluational approaches. Second, Paul Horwich's epistemicist proposal is examined with a special focus on the treatment of the Liar paradox. Horwich's account about how to construct his theory of truth is formalized and critically discussed with the use of a fixed-point construction. In the last chapter, I introduce and discuss some logics based on the work of Hartry Field that use two conditionals in a language with a truth predicate and vague predicates.

RESUM

En aquesta tesi examino algunes de les propostes més importants de solució comuna a les paradoxes del Mentider i la *Sorites*. Per tal de fer-ho, introdueixo, primer, una definició de la noció de paradoxa i, amb ella, caracteritzo en detall què cal esperar d'una solució comuna a un grup de paradoxes. A continuació, considero quines són les raons que podem tenir per tal d'adoptar una solució comuna a una colecció de paradoxes i extrec algunes conclusions respecte les propostes de Vann McGee i Graham Priest per fer front al Mentider i la *Sorites*. En els tres capítols següents, examino tres autors en detall. Primer, rebutjo la proposta de Jamie Tappenden per inapropiada, especialment en el cas del Mentider. Pel que fa a la *Sorites*, mostro que la teoria que Tappenden defensa és, al menys, tan problemàtica com les propostes superavaluacionistes. En segon lloc, examino la teoria epistemicista de Paul Horwich, amb especial atenció a la seva aplicació al mentider. A través d'una construcció de punt fixe, formalitzo i discuteixo críticament la proposta de Horwich sobre com construir la seva teoria de la veritat. En l'últim capítol, introdueixo i discuteixo algunes lògiques, basades en les propostes de Hartry Field, que usen dos condicionals en llenguatges amb un predicat de veritat i predicats vagues.

ACKNOWLEDGEMENTS

First of all, I would like to express my deepest gratitude to my supervisor, José Martínez, for his excellent guidance, support, patience and for providing me with the best atmosphere for doing research.

I would also like to express my gratitude to the members of my thesis committee: Dan López de Sa, Gabriel Uzquiano and Elia Zardini. I would like to thank them for accepting to critically read my thesis.

I would also like to express my sincere gratitude to my tutor, Manuel García-Carpintero.

I had the opportunity to begin my PhD thanks to a FPU scholarship from the Spanish Ministry of Education and Science. I also enjoyed two visiting research grants from the same institution; the first at the University of Sheffield and the second at the New York University. This dissertation has been possible thanks to a number of research projects: *Reference, Self-Reference and Empirical Data* (FFI2011-25626) from the Spanish Ministry of Education; *PERSP—Philosophy of Perspectival Thoughts and Facts* (CSD2009-00056) from the Spanish Ministry of Education and Science; *Localism and Globalism in Logics and Semantics* (FFI2015-70707-P) from the Spanish Ministry of Economy and Competitiveness.

I would like to thank all the people that, at one moment or another, were at *Logos* during the time I was working on this dissertation, specially Dan López de Sa, Manuel Martínez, Manuel Pérez Otero, Sven Rosenkranz and Elia Zardini for helping me with invaluable comments to earlier drafts of this thesis.

At the University of Barcelona, I would also like to express my gratitude to Calixto Badesa and Joost Joosten for their comments and support.

During these years at Barcelona I have made many friends, who helped me in many ways: Marc Artiga, Joan Bertran, María Esteban, Marta Jorba, Paco Murcia, Chiara Panizza,Umberto Rivieccio, Gonçalo Santos, Luz Sabina, Pilar Terrés and, of course, Oscar Cabaco, whose friendliness and help has

been invaluable.

At Sheffield, I would like to thank, first, my supervisor, Rosanna Keefe, who helped me and provided me with extremely helpful discussions. Also, at Sheffield, I want to thank Kathy Puddifoot, Kate Harrington, Paniel Reyes, Bernardo Aguilera, Inga Vermeulen, Julien Murzi and Jonathan Payne. I would like to specially thank, at Barcelona, London Ontario, Sheffield and everywhere, Cristina Roadevin.

At New York, I am grateful to my supervisor, Hartry Field, for sharing with me some of his time and providing me with extremely helpful discussions. I would also like to thank Paul Horwich, for his patience and help, and David Samsundar, for his friendliness and warm farewell.

Per acabar, vull agrair a tota la meva família l'esforç que, durant aquests anys, han fet per ajudar-me a acabar aquest treball. En especial, a l'Oriol Oms, per animar-me a acabar; al Josep Maria Sardans, per les seves preguntes i, sobretot, als meus pares, el Ramon i la Magda, per la seva paciència. Deixo, també, un record i tot el meu agraïment pel Pepito i la Lluïsa.

Però el meu agraïment més profund és per la Cristina, sense la qual aquesta tesi no hauria estat possible, i pel Guillem i l'Alícia, sense els quals aquesta tesi s'hauria acabat tres anys abans.

CHAPTER

ONE

INTRODUCTION

1.1 Truth

The investigation on the notion of truth is one of the main problems in philosophy and also one of the oldest ones. At least since Pilate asked 'what is truth?' (*John*, 18:38), human beings have tried to look into the nature of truth.

As Russell (1950) sensibly stated, looking at the question 'What is truth?' —so as to obtain a *general definition of truth*— is not the same as asking 'Which beliefs, or sentences, are true?' —so as to obtain a *general criterion for truth*. But yet, a clear answer to the former might be an invaluable help to respond to the latter. Moreover, since often enough philosophical problems have their roots in confusion rather than in ignorance, a theory that clarifies the concept of truth makes easier any approach to theories containing such concept, for it elucidates one of the sources of confusion.

The inquiry into the nature of truth has bred many different proposals. Among the main ones it is worth mentioning the following.[1]

The correspondence theory of truth Inspired by Aristotle, who cryptically defended it in his *Metaphysics* (Aristotle (1984, 1011b25)) and, later, by the work of Thomas Aquinas (Aquinas (1981, I^a, Q. 16)), this is very likely the most venerable of the theories of truth. It defends that being true consists in corresponding to facts. Different versions of the

[1] For an overview of the many theories of truth present in the philosophical landscape see, for example, Mackie (1973), Kirkham (1995), Lynch (2001), Künne (2003), Blackburn and Simmons (2005), Burgess and Burgess (2010) or Glanzberg (2014).

correspondence theory of truth have been defended in, for example, Russell (1950) and Austin (1950).

The coherence theory of truth According to the coherence view on truth, being true consists in being coherent with a certain given collection of other truth-bearers. It has been defended, in different forms, in Bradley (1914) or, more recently, in Young (1995).

The pragmatic theory of truth Pragmatists about truth think that being true consists in being useful in practice. The pragmatic theory of truth has been defended, specially, in the work of the American philosopher William James (see, for instance, James (1907)).

The deflationary theory of truth According to Dummett (1959) this view originated with the work of the German philosopher Gottlob Frege (see, for instance, his 1918). The deflationary theory of truth defends the view that truth is not a genuine, or robust, property, but a deflationary one; this means that the truth predicate is not used to describe anything, in the sense that truths do not share any interesting common property. Some forms of Deflationism can even stress this point a little further and defend that truth is not a property at all.[2]

Deflationists think that asserting that something is true is equivalent to asserting this something itself, although, of course, the nature of such equivalence may vary from one philosopher to another. Moreover, nothing more is needed beyond this equivalence to explain all facts concerning truth. Two of the main contemporary defenders of deflationism are Paul Horwich and Hartry Field (see Horwich (1998b) and Field (1994)). Both will be discussed in this dissertation; the former in Chapter 6 and the latter in Chapter 7.

Any investigation into the nature of truth sooner or later has to face one of the oldest and toughest paradoxes in Philosophy of logic and language: the Liar paradox. In the following section I will introduce such paradox together with some other paradoxes closely related to it.

[2]Dorothy Grover, for example, claims that the expressions 'true' and 'false' are prosentences; that is, like pronouns, they are expressions without operative meaning unless they inherit it from other expressions. Grover has defended the Prosentential account, for instance, in Grover (1992) or Grover (2001)

1.2 The Liar

Suppose we have a sentence λ that asserts its own untruth (I will call such a sentence *the Liar sentence* or, sometimes, just *the Liar*):

(λ) λ is not true.

Suppose, now, that λ is true. Then, if λ is true what λ says is the case (is not that what truth is about?); but what λ says is precisely that λ is not true, which contradicts our initial supposition that λ is true. Thus, we confidently conclude, by *reductio* that λ is not true. Unfortunately, if λ is not true, then what λ says (that is, that λ is not true) is the case and, thus, λ is true after all, which contradicts our previous conclusion. We suddenly realize, then, with astonishment that we are stuck.

This is an informal presentation of the Liar paradox. The first version of this paradox is often credited to Eubulides of Miletus (fourth century BC), contemporary of Aristotle, and famous for his seven puzzles: the Liar, the Disguised, Electra, the Veiled Figure, The Sorites, The Horned One and the Bald Head (Laertius 1972, p. 237). There is still another purported version of the paradox, though, attributed to the Cretan philosopher Epimenides (sixth century BC); according to St. Paul's Epistle to Titus, Epimenides said that all Cretans were liars (*Titus*, 1:12).

Notice that Epimenides' assertion is not necessarily paradoxical; actually, in most ordinary situations it will be plainly false (for example, if there is some Cretan who is not a liar). Its paradoxicality only arises in certain situations that can be easily devised. Suppose, first, that a liar is a person who has never (even unintentionally) said anything true. Consider, second, a situation where all cretan utterances are false (except perhaps Epimenides' one). Then, in this case, if the sentence 'all Cretans are liars' is true, this very same sentence is false, for it has to be a lie (since it has been uttered by a Cretan). But if it is false, then it is a lie and, consequently (given that the rest of Cretan utterances are false) it is true. The distinction between Eubulides' and Epimenides' version of the Liar brings up one important feature of this paradox: some sentences (like Eubulides' one) are intrinsically paradoxical in the sense that they can generate a paradox independently of the way the world is; they are paradoxical in all possible worlds. On the other hand, some sentences (like Epimenides' one) generate paradoxes that do not depend only on semantic stipulations, but also on empirical facts in the world.

Kripke (1975) remarks precisely this distinction when he shows the following way of achieving self-reference. Suppose P and Q are two predicates of a given language that apply to sentences of that very language. Then the sentence

$$\forall x(Px \rightarrow \neg Qx)$$

can be paradoxical if some empirical facts determine that this very same sentence is the only one that satisfies P and, moreover, Q is the truth predicate. For instance, suppose that the predicate P is the predicate 'the sentence written on the blackboard of room 202'. Now, depending on the way things are, the sentence $\forall x(Px \rightarrow \neg Qx)$ can be paradoxical; it is only needed that this very same sentence, and only it, is written on the blackboard of room 202 and that the predicate Q is interpreted as the truth predicate. That means that utterances in which we use the truth predicate might be paradoxical without our being aware of it. Following the habitual usage, I will call paradoxes like this *contingent Liar paradoxes* and the sentences that generate them *contingent Liar sentences*.

Apart from the contingent Liar paradox, there is another liar-like paradox which is worth mentioning: the Liar Cycles paradox.[3] Let me present a version of such a paradox. Consider these three sentences:

(λ_1) λ_2 is true.

(λ_2) λ_3 is true.

(λ_3) λ_1 is not true.

We can now informally reason as follows. Suppose λ_1 is true. Then what it says is the case and, hence, λ_2 is true, what, in turn, implies that λ_3 is true, what, finally, lets us conclude that λ_1 is not true (for that is what λ_2 says). So, from the supposition that λ_1 is true, we reached a contradiction and, hence, we can conclude that λ_1 is not true. But if λ_1 is not true, then λ_3 is true (for λ_3 says, precisely, that λ_1 is not true) and, hence, λ_2 is true and, consequently, λ_1 is true, contradicting our previous conclusion.

It is easy to see that the Liar Cycles paradox can be constructed with any finite cycle of sentences we like. Another preeminent feature of this paradox is that it shows that simple self-reference is not necessary for having a Liar-like paradox, but mere circularity is enough; even if in the last example above there is not a sentence who directly claims its own untruth, it seems clear that sentence λ_3 is indirectly claiming its own untruth via the other two sentences.

[3]This version of the Liar paradox is sometimes called *Jourdain's paradox*, after the British logician Philip Jourdain who, the myth says, was the first to introduce it. Its version with just two sentences is usually called *the Card paradox*.

It is worth introducing another Liar-like paradox, related to the Liar Cycles paradox, intended to show that, in order to have paradoxical results, not even circularity is necessary: Yablo's paradox.

Yablo (1993) claims that his paradox shows that a Liar-like paradox can be developed even in the absence of self-reference or circularity. The paradox is the following one.

Suppose you have an infinite sequence of sentences such that any sentence claims that every subsequent sentence is not true:

(S_0) for all $k > 0$, S_k is not true,

(S_1) for all $k > 1$, S_k is not true,

(S_2) for all $k > 2$, S_k is not true,

$\vdots$

(S_n) for all $k > n$, S_k is not true,

$\vdots$

Now, we can reason informally as follows. For any n:

1. if $S_n$is true, then for all $k > n$, S_k is not true,
2. that means that, in particular, S_{n+1} is not true.
3. 1 also implies that for all $k > n + 1$, S_k is not true,
4. but then, S_{n+1} is true; contradiction.
5. Since supposing $S_n$is true leads to a contradiction, we must conclude that S_n is not true.
6. since n was arbitrary, we conclude, by universal generalization, that for all k, S_k is not true.
7. Now we can end the argument by showing that, by 6, for all $k > 0$, S_k is not true,
8. which implies that S_0 is true; contradiction (for we showed that all sentences were not true).

If we really have a Liar-like paradox without circularity, then we have shown that circularity is not a necessary condition for having a paradoxical result. However, there is no consensus on whether Yablo's paradox involves some kind of circularity; Graham Priest, for instance, has defended that the kind of paradox just sketched does involve self-referential circularity (see Priest 1997 or Beall 2001) and, besides Yablo himself (in the aforemetioned Yablo 1993 and also in Yablo 2004), other authors have defended that this paradox is a genuine case of a paradox not involving circularity (see, for example, Sorensen 1998).

I will finish this section with another semantic paradox closely related to the Liar: Curry's paradox.[4] Let γ be a sentence, called a *Curry sentence*, which asserts that if itself is true then ϕ, where ϕ is any sentence whatsoever:

(γ) If γ is true, then ϕ.

Suppose, now, that γ is true. Then, under this supposition, what γ says is the case; that is, if γ is true, then ϕ. We can now apply *Modus Ponens* and conclude, under the assumption that γ is true, ϕ. Since we have achieved ϕ under the supposition that γ is true we can conclude, now, that *if* γ is true, *then* ϕ. We realize, now, that we just proved γ itself. But if γ is the case —as we have just proven— then γ is true. This means that we have, now, the following two claims: first, that if γ is true, then ϕ and, second, that γ is true. Finally, then, applying *Modus Ponens* again, we astonishingly conclude ϕ. But ϕ was any sentence, in particular, thus, it could be a contradiction or any false sentence (of course, though, it could also be a true sentence).

1.3 The Formal Liar

We ought to ask ourselves, now, what properties should be expected from a truth predicate. As we will see, what characterizes truth is, above all, the close connection there is between a sentence and its truth ascription.

Following this line of thought, a highly desirable property of a truth predicate Tr is the so called *Intersubstitutivity Principle*:[5]

[4] This paradox is named after Haskell B. Curry who first presented it in his (1942). Some authors, given the similarity of the reasoning used in the paradox with the reasoning used in the proof of Löb's theorem have called this paradox *Löb's paradox* (see, for instance, Barwise and Etchemendy 1984).

[5] All the desirable properties for the truth predicate in this context are implicitly stated for non opaque contexts.

(IP) If two sentences ϕ and ψ are alike except that one has a sentence χ where the other has $Tr^{\ulcorner}\chi^{\urcorner}$, then $\phi \vdash \psi$ and $\psi \vdash \phi$.[6]

Another two features we might want our truth predicate to validate are the following ones. First, an *ascent* principle or *truth introduction* principle; for any sentence ϕ:

(Tr_i) $\phi \vdash Tr^{\ulcorner}\phi^{\urcorner}$.

And second, a *descent* principle or *truth elimination* principle; for any sentence ϕ:

(Tr_e) $Tr^{\ulcorner}\phi^{\urcorner} \vdash \phi$.

Finally, another characteristic we might expect our truth predicate to have is expressed by the *T-schema*, so called after Alfred Tarski; for any sentence ϕ:

(T-schema) $\phi \leftrightarrow Tr^{\ulcorner}\phi^{\urcorner}$

All these principles are closely related and all of them, as I said, try to capture the close relation there is between a sentence and its truth ascription. Notice that, if a sentence implies itself, (IP) clearly entails (Tr_i) and (Tr_e). The other way around, though, is not the case; we may have a theory of truth which validates (Tr_i) and (Tr_e) but which fails to validate (IP) and, hence, fails to validate the principle of intersubstitutivity of logical equivalents.

The T-schema depends on the properties of the conditional. If the conditional validates the principle '$\phi \rightarrow \phi$' for any sentence ϕ, then (IP) will imply the T-schema and if the logic has the principle of conditional proof (if $\phi \vdash \psi$ then $\phi \rightarrow \psi$ is a logical truth), (Tr_i) and (Tr_e) will imply the T-schema. In classical logic all such principles are equivalent and, thus, Tarski captured all of them when he presented the T-schema (see, for example, Tarski 1944, 1983). According to him, the T-schema is a minimal material condition for any theory of truth; the instances of the T-schema should be implied by any characterization of the notion of truth (Tarski 1944, p. 344).

In order to arrive to a contradiction as a conclusion of the Liar paradox in classical logic we need, on the one hand, the intuitive properties of the truth predicate captured by the principles above and, on the other hand, we also need a sentence that somehow *says* of itself that it is not true; we can

[6]Here, the brackets ('$\ulcorner$' and '$\urcorner$') indicate some device of canonical name formation for sentences; thus, '$\ulcorner\phi\urcorner$' is just a name for the sentence ϕ and '$\mathrm{Tr}^{\ulcorner}\phi^{\urcorner}$' ascribes truth to the sentence ϕ.

construct such a sentence either with the use of plain self-reference (with a sentence mentioning itself) or some kind of circularity or, finally, certain surrogates of self-reference that are enough to create the paradox —in particular, we will show how to have a sentence that mentions another sentence that is equivalent to the first one. Gödel showed how to construct such a sentence in what is usually called the *Diagonal Lemma*.

In order to sate and prove the Lemma, let us suppose we have a classical first-order language $\mathcal{L}^+ = \mathcal{L} \cup \{Tr\}$ with a certain monadic predicate Tr and suppose, furthermore, that for every formula $\phi \in \mathcal{L}^+$ we can express its canonical name $\ulcorner \phi \urcorner$ in $\mathcal{L}$ via some Gödel codification. Gödel (1931) showed that, if $\boldsymbol{T}$ is a theory that can be appropriately axiomatized and contains Robinson's arithmetic, which is weaker than Peano's arithmetic, then we can have canonical names for all the expressions in the language, we can express the syntax of the language and we can prove the following important Lemma.[7]

Lemma 1.3.1 (Diagonal Lemma) *If $\phi(x)$ is a formula of $\mathcal{L}^+$ with one free variable, then there is a sentence δ of $\mathcal{L}^+$ such that $\boldsymbol{T} \vdash \delta \leftrightarrow \phi(\ulcorner \delta \urcorner)$.*

Proof Let us define, first, the *diagonal function*, $d(x)$, which assigns to any formula $\phi(x)$ of $\mathcal{L}^+$ with one free variable the sentence which is the result of substituting the canonical name of ϕ for x in $\phi(x)$; that is, $d(\phi(x)) = \phi(\ulcorner \phi(x) \urcorner)$. That a sentence y is the diagonalization of some formula x can then be expressed by a formula of $\mathcal{L}^+$, $diag(x, y)$. Now consider the following sentence:

$$\psi(y) =_{def} \forall z(diag(y, z) \rightarrow \phi(z))$$

which is a sentence with one free variable stating that if a given sentence is the diagonalization of y, then that sentence, so to speak, has the property expressed by ϕ.

We can now diagonalize $\psi(y)$ so that its diagonalization is δ; that is, $diag(\ulcorner \psi(y) \urcorner, \ulcorner \delta \urcorner)$. Notice now that, if δ is the diagonalization of $\psi(y)$, that means that, by definition of the diagonal function, $\delta \leftrightarrow \psi(\ulcorner \psi \urcorner)$ and, hence, $\delta \leftrightarrow \forall z(diag(\ulcorner \psi \urcorner, z) \rightarrow \phi(z))$.

Again, now, if $diag(\ulcorner \psi(y) \urcorner, \ulcorner \delta \urcorner)$, then $\forall z(diag(\ulcorner \psi(y) \urcorner, z) \leftrightarrow z = \ulcorner \delta \urcorner)$. This last equality implies, hence, $\forall z(diag(\ulcorner \psi(y) \urcorner, z) \rightarrow \phi(z)) \leftrightarrow \forall z(z =$

[7] I am simplifying the presentation of the Diagonal Lemma; I am specially loose with the use-mention distinction. For details of a modern formulation see, for example, Boolos, Burgess, and Jeffrey (2002) or Smith (2007). The proof here is adapted from the latter.

$\ulcorner\delta\urcorner \rightarrow \phi(z)$), because the antecedents of the conditionals at each side of the biconditionals have been proved to be equivalent.

But we have just seen that the left-hand side of this last biconditional is equivalent to δ and the right-hand side is just stating that $\phi(\ulcorner\delta\urcorner)$. Hence, we conclude $\delta \leftrightarrow \phi(\ulcorner\delta\urcorner)$. □

Once we have the Diagonal Lemma it is simple to formulate the Liar paradox. Suppose we are in $\boldsymbol{T}$ as before and that we interpret the predicate Tr as a truth predicate (that is, $\boldsymbol{T}$ contains some axioms that make Tr into a truth predicate for $\mathcal{L}^+$). That means that Tr obeys the T-schema (as it was said before, if we are in classical logic this is just equivalent to (IP) on the one hand and, on the other, to (Tr_i) and (Tr_e) together). That is, for any sentence $\phi \in \mathcal{L}^+$, $\boldsymbol{T} \vdash Tr\ulcorner\phi\urcorner \leftrightarrow \phi$. But now, using the Diagonal Lemma and taking $\neg Tr(x)$ as $\phi(x)$, we obtain a sentence $\lambda \in \mathcal{L}^+$ such that $\boldsymbol{T} \vdash \lambda \leftrightarrow \neg Tr\ulcorner\lambda\urcorner$. It is clear now that this last biconditional together with the λ-instance of the T-schema, $Tr\ulcorner\lambda\urcorner \leftrightarrow \lambda$, yield $Tr\ulcorner\lambda\urcorner \leftrightarrow \neg Tr\ulcorner\lambda\urcorner$, which, in classical logic is equivalent to $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$. We then conclude that $\boldsymbol{T} \vdash Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$.

It turns out that, in order to reach a contradiction, the full biconditional of the T-schema is not needed. Montague (1963) showed that it is enough to have one direction of it together with the rule version of the other direction. Thus, suppose we have now an extended language $\mathcal{L}^+$ with a new predicate τ and an axiomatic theory $\boldsymbol{S}$ such that, for any sentence $\phi \in \mathcal{L}^+$:

(I) $\boldsymbol{S} \vdash \tau\ulcorner\phi\urcorner \rightarrow \phi$

(II) If $\boldsymbol{S} \vdash \phi$ then $\boldsymbol{S} \vdash \tau\ulcorner\phi\urcorner$

Then we reason as follows, using, again, the Liar sentence:

1.	$\neg\tau\ulcorner\lambda\urcorner \rightarrow \lambda$	(Diagonal Lemma)
2.	$\tau\ulcorner\lambda\urcorner \rightarrow \lambda$	(I)
3.	λ	(1, 2 and logic)
4.	$\tau\ulcorner\lambda\urcorner$	(II on 3)
5.	$\tau\ulcorner\lambda\urcorner \rightarrow \neg\lambda$	(Diagonal Lemma)
6.	$\neg\lambda$	(*Modus Ponens* on 4 and 5)
7.	$\lambda \wedge \neg\lambda$	(3 and 5)

Principles (I) and (II) can be reasonably expected of notions like knowledge and necessity (besides, of course, truth itself). Thus, Montague's argument shows that paradoxicality dangerously spreads from truth to other related notions.

We can also formulate in a precise way Curry's paradox, where the Diagonal Lemma is used to obtain a sentence $\gamma \in \mathcal{L}^+$ such that $\gamma \leftrightarrow (Tr\ulcorner\gamma\urcorner \rightarrow \phi)$, where ϕ is any sentence of $\mathcal{L}^+$. Let us, then, reason as follows:

1. $\gamma \leftrightarrow (Tr\ulcorner\gamma\urcorner \rightarrow \phi)$	(Diagonal Lemma)
2. $Tr\ulcorner\gamma\urcorner \rightarrow (Tr\ulcorner\gamma\urcorner \rightarrow \phi)$	(γ-instance of the T-schema and logic on 1)
3. $Tr\ulcorner\gamma\urcorner \rightarrow \phi$	(contraction —$\psi \rightarrow (\psi \rightarrow \chi) \vdash \psi \rightarrow \chi$— on 2)
4. $(Tr\ulcorner\gamma\urcorner \rightarrow \phi) \rightarrow Tr\ulcorner\gamma\urcorner$	(γ-instance of the T-schema and logic on 1)
5. $Tr\ulcorner\gamma\urcorner$	(*Modus Ponens* on 3 and 4)
6. ϕ	(*Modus Ponens* on 3 and 5)

1.4 The *Sorites*

Another of the logical puzzles that Eubulides of Miletus held was the *Sorites* paradox. The Greek word 'sorites' is an adjective cognate with the noun 'soros', which means heap. In antiquity this kind of puzzles were usually presented in the form of a series of questions; thus, in the framework of the disputes between the empirical doctors and the dogmatic doctors, Galen says in his *On Medical Experience*:

> I say: tell me, do you think that a single grain of wheat is a heap? Thereupon you say: No. Then I say: What do you say about two grains? For it is my purpose to ask you questions in succession, and if you do not admit that 2 grains are a heap then I shall ask you about three grains. Then I shall proceed to interrogate you further with respect to four grains, then five and six and seven and eight, and you will assuredly say that none of these makes a heap.[...] I for my part shall not cease from continuing to add one to the number in like manner [...]. It is not possible for you to say with regard to any one of these numbers that it constitutes a heap. [...] If you do not say with respect to any of the numbers [...] that it now constitutes a heap, but afterwards when a grain is added to it, you say that a heap has now been formed, consequently this quantity of corn becomes a heap by the addition of the single grain of wheat, and if the grain is taken away the heap is eliminated. (Galen, *Medical Experience*, XVI 2, cited in Barnes 1982, p. 33)

Galen concludes that it is absurd to claim that a single grain determines the existence or inexistence of a heap.[8] Diogenes Laertius does not claim

[8]Actually, this is essentially what today is known as *the Forced-March Sorites*. Suppose we are in front of a *soritical* series like, for example, a series of collections

that Eubulides is the inventor of these arguments; as a matter of fact, some authors see the origins of the *Sorites* paradox in Zeno's paradox of the Millet Seed, according to which, since a bushel of seed makes a noise as it falls to the ground, any individual seed must also make a noise on falling; the thing is, though, that since Zeno's reasoning was based on principles of proportionality rather than on soritical reasoning, it is not very likely that the Millet Seed was the origin of the *Sorites*.

In the contemporary discussion on the *Sorites*, it is presented in the form of an argument:

1. A man with no hairs on his head is bald.

2. If a man with n hairs on his head is bald then a man with $n+1$ heads on his head is bald.

3. Therefore, a man with a million hairs on his head is bald.

As we will see in Section 1.5, the *Sorites* paradox can have many different formulations. It can be already noted, though, that the argument above needs only a little amount of logical resources; the only rules that it requires are universal instantiation and *Modus Ponens*.

The possibility of constructing this paradox is one of the preeminent features of the vague predicates. More specifically, what is in the core of the *Sorites* is what Wright (1975, p. 333) calls *tolerance*. We can find in vague predicates a certain tolerance with respect to their application to objects which differ only in some small changes in the relevant aspects. For instance, in the example presented above, small differences in the number of grains do not make any difference to the application of the predicate 'heap'. The paradox arises because large changes in the relevant aspects *do* make a difference to the application of vague predicates, large changes that can be achieved via

---

of grains of wheat such that each member of the collection has just one grain more than the previous one, and suppose that the series begins with a single grain (a clear case of not being a heap) and ends with a collection of 10^6 grains of wheat (a clear case of being a heap). In a Forced-March Sorites we are asked to imagine an idealized subject S who is prompted to answer the question 'is this collection of grains of wheat a heap?' for each of the members of the series in ascending order. Then, S's first answer will surely be 'No' (or something equivalent) and we would expect of S to stick to this answer, since a grain of wheat does not make any difference for being a heap or not. But there must be a point in which S answer has to change, unless she wants to be answering 'No' to the question 'is a collection of 10^6 grains of wheat a heap?'. The term was first introduced in Horgan (1994); see also Keefe (2000, p. 25).

a big number of small ones. Tolerance can be represented by some *tolerance relations* which would relate the objects that differ very little with respect to the relevant aspects used in the formulation of the paradox; thus, for example, in the case of the predicate 'heap', the tolerance relation, call it R, we used in Galen's example is, for any collections of grains of wheat a and b:

Rab iff b has one grain more of wheat than a

Another one of the main features of vague expressions is the existence of borderline cases; cases in which it is not clear if the term applies or not. Thus, for example, it seems that no conceptual analysis nor any empirical investigation can settle whether a 1.8 meter man is tall or not. Thus, for instance, Max Black, in his (1937), defines the vagueness of a predicate as "the existence of objects concerning which it is intrinsically impossible to say either that the [predicate] in question does, or does not, apply" (Black 1937, p. 430). Saying that it is intrinsically unclear whether a given predicate applies or not to an object means that if we cannot determine if the predicate applies or not to the object is due to the very semantic nature of the predicate, not to some external condition. Horwich, as we will see in chapter 6, similarly thinks that the existence of borderline cases is the main feature of vague predicates; according to him, in front of a given ascription of a vague predicate F to an object which is a borderline of F we are not disposed to accept such ascription nor its negation and, moreover, we do not think further investigation will help solve the matter.

Finally, vague predicates also lack sharp boundaries; not only between the clear cases and the clear countercases, but also between the clear cases and the borderline cases, and between the clear cases and the borderline cases of the borderline cases, and so on. This phenomenon is known as *higher-order vagueness* and is another of the main features of vague expressions.

So we can see that the existence of higher order vagueness, borderline cases, tolerance and the *Sorites* paradox gives a precise enough idea of what a vague predicate is.

1.5 Forms of the *Sorites*

The *Sorites* paradox comes in many ways.[9] We can follow Barnes (1982, p. 30) and notice that, in order to construct a paradox of vagueness with a given vague predicate P, it is sufficient to have an ordered series of objects $a_1, a_2, \ldots a_n$ such that Pa_1, $\neg Pa_n$ and such that all adjacent objects in the

[9] In this section I am loosely following Hyde (2011, 2014).

series must be related by the tolerance relation; that is, for each $0 \leq i < n$, a_i and a_{i+1} must be indiscriminable with respect to the application of P; in terms of the tolerance relation R for P, for each $0 \leq i < n$, Ra_ia_{i+1}. Although these conditions will do for now, in chapter 2 I will defend that they are sufficient, but not necessary.

We already mentioned the form of the paradox that uses just *Modus Ponens*; for a given vague predicate P and a series of objects $a_1, a_2, \ldots a_n$ satisfying the conditions stated in the previous paragraph the following is an schematic representation of what is sometimes called *the Conditional Sorites*:

Conditional *Sorites*

1 Pa_1

2 If Pa_1 then Pa_2

3 Hence, Pa_2

4 If Pa_2, then Pa_3

5 Hence, Pa_3

$\vdots$

$2n-2$ If Pa_{n-1}, then Pa_n

$2n-1$ Hence, Pa_n.

In this formulation we are presupposing that the adjacent objects in the soritical series satisfy the tolerance relation, which could have been made explicit. This version of the paradox can also be formulated with a biconditional instead of a conditional (clearly, the right to left directions of the conditionals above are uncontroversial). Following this, Priest (2010a, p. 73) defends that the *Sorites* paradox formulated with the material biconditional is the best way to capture the tolerance of vague predicates; to capture tolerance, according to Priest, we need the ascriptions of the predicate to adjacent objects in the series to have the same truth value. And there is nothing more to tolerance than that.[10]

Dummett (1975) presents another version of the paradox that is specially virulent; we can formulate the paradox using phenomenological predicates like, for instance, 'looks red'. Take, for example, a *soritical* series of patches of colors indistinguishable between them such that begins with a clearly red

[10] We will come back to this issue in chapter 4, section 4.2.

patch (a_1) and ends with a clearly yellow patch (a_n). In this case, we can consider the relation

Rxy iff x and y look the same color

which, in this case, makes reasonable to allow substitution of x and y in the predicate 'looks red' when Rxy and, thus, we can formulate the paradox without any appeal to conditionals:

Phenomenological *Sorites*

1 a_1 looks red

2 Ra_1a_2

3 Ra_2a_3

$\vdots$

n $Ra_{n-1}a_n$

n+1 Hence, a_n looks red.

In the same line, we can also formulate the paradox using identity (see Priest 1991, 2010b). Suppose we have a color *soritical* series as in the previous example. Let b_i be 'the phenomenological look of a_i' (for $1 \leq i \leq n$). Then, in this case, using transitivity of identity and the fact that, for each $1 \leq i \leq n$, $b_i = b_{i+1}$, we can easily conclude that $b_1 = b_n$; that is, that the phenomenological look of a clearly red patch is equal to the phenomenological look of a clearly yellow patch.

The paradox can also be formulated using the idea that vague predicates do not determine sharp boundaries between the objects to which they apply and the objects to which they do not apply:

No-Sharp-Boundaries *Sorites*

1 Pa_1

2 $\neg(Pa_1 \wedge \neg Pa_2)$

3 Hence, Pa_2

4 $\neg(Pa_2 \wedge \neg Pa_3)$

5 Hence, Pa_3

$\vdots$

$2n-2$ $\neg(Pa_{n-1} \wedge \neg Pa_n)$

$2n-1$ Hence, Pa_n.

Notice that, in classical logic, the Conditional *Sorites* and the No-Sharp-Boundaries *Sorites* use equivalent sentences, but that they not need be equivalent in other logics.

We can use some more logical resources, in particular, universal instantiation:

Induction *Sorites*

1 Pa_1

2 $\forall x(Pa_x \rightarrow Pa_{x+1})$

3 Hence, Pa_n

The same can be schematically represented with an explicit use of the tolerance relation:

1 Pa_1

2 $\forall x((Pa_x \wedge Ra_x a_{x+1}) \rightarrow Pa_{x+1})$

3 Hence, Pa_n

The second premise is usually called *the inductive premise*. In the same way that the Induction *Sorites* relates to the Conditional *Sorites*, we can formulate the No-Sharp-Boundaries paradox using the existential quantifier:

∃-No-sharp-boundaries *Sorites*

1 Pa_1

2 $\neg\exists x(Pa_x \wedge \neg Pa_{x+1})$

3 Hence, Pa_n

Again, we could have explicitly used the tolerance relation.

We can also present the same paradox in a way that stresses the pressure that the *Sorites* paradox puts on accepting the existence of sharp boundaries between the extensions (that is, the objects to which the predicate applies) and the anti-extensions (that is, the objects to which the predicate does not apply) of vague predicates:

Line-drawing *Sorites*

1 Pa_1

2 $\neg Pa_n$ $(n > 1)$

3 $\forall x(Pa_x \vee \neg Pa_x)$

4 Hence, there is a z $(1 \leq z < n)$ such that Pa_z and $\neg Pa_{z+1}$

In the reasoning underlying this formulation the Least Number Principle is used[11] (which, in classical logic, is equivalent to the Induction Principle).

Traditionally, the small changes used to characterize tolerance are discrete. This does not mean, though, that a continuous version of the paradox cannot be stated. Weber and Colyvan (2010) present a continuous version of the *Sorites* paradox, which I am going to sketch in the following.

Given a vague predicate P, we can consider the real-number interval $[0,1]$ partitioned in the following way:

$$A = \{x \in [0,1] | P(x)\}$$
$$B = \{x \in [0,1] | \neg P(x)\}$$

with the following restrictions:

(i) A and B are non-empty

(ii) $0 \in A$,

(iii) $1 \in B$,

(iv) for each $x \in A$ and each $y \in B$, $x < y$,

(v) for each $x, y \in [0,1]$, if $x < y$ and $\neg P(x)$, then $\neg P(y)$ (and its contrapositive).

That is, A is the left-hand side of the interval $[0,1]$ and B is its right-hand side. The properties of $\mathbb{R}$ guarantee that A has a least upper bound in $[0,1]$, call it $\sup A$ and also that B has a greatest lower bound in $[0,1]$, call it $\inf B$. Now, tolerance of P allows us to conclude that $P(\sup A)$. This is so

[11] The least number principle states that every non-empty set of natural numbers has a least element. That is, if the range of the quantifiers is the set of natural numbers $\mathbb{N}$ and Φ is a formula that expresses a given subset of $\mathbb{N}$, then the Least Number Principle claims that $\exists n \Phi(n) \rightarrow \exists n(\Phi(n) \wedge \forall x(x < n \rightarrow \neg\Phi(x)))$.

by what Priest (2006) calls *the Leibniz continuity condition*; the idea is that if something is going on as close as we wish to a given limit point, then it is also going on at the limit point. Of course this is not valid generally, but it can be taken to be a generalization of tolerance applied to discrete cases. For the same reason, we can conclude $\neg P(\inf B)$. Now we can reason as follows.

If $\sup A \neq \inf B$, then, by the fact that $\mathbb{R}$ is linearly ordered, we either have that $\sup A < \inf B$ or $\inf B < \sup A$. Given that the reals are dense (that is if $x, y \in \mathbb{R}$ are such that $x < y$, then there is a $z \in \mathbb{R}$ such that $x < z < y$) there will be either a $z_1 \in \mathbb{R}$ such that $\sup A < z_1 < \inf B$, or a $z_2 \in \mathbb{R}$ such that $\inf B < z_2 < \sup A$. But notice that, by definition, for each x, if either $x > \sup A$ or $x > \inf B$, then $\neg P(x)$ and, for each x, if either $x < \inf B$ or $x < \sup A$, then $P(x)$. It follows then that $P(z_i)$ and $\neg P(z_i)$ $(1 \leq i \leq 2)$.

Hence, $\sup A = \inf B$. But then, since, by tolerance, $P(\sup A)$ and $\neg P(\inf B)$, we conclude $P(\sup A)$ and $\neg P(\sup A)$. Contradiction.

1.6 Vagueness

Some, if not most, of the predicates we use every day in natural languages are vague; expressions like 'bald', 'rich', 'tall', 'red',... are vague predicates in the sense specified in section 1.4. Among the theories of vagueness that have been proposed in the endeavor to clarify such phenomenon and solve the *Sorites* paradox it is worth to take into account the following.[12]

Supervaluationist approaches Supervaluationists think the origin of vagueness is in the language and they manage to retain classical logic while accepting truth-value gaps; they can achieve that with the use of Van Fraassen's (see his 1966) semantics, which is not truth-functional, applied to vagueness. Supervaluationists defend that the second premise in the Induction *Sorites* is false and, hence, the argument is not sound; equivalent strategies can be found with respect to the other verions of the paradox. The view originated with Fine (1975) and has been defended, among others, in Keefe (2000).

Many-valued approaches These theories, like Supervaluationist accounts, see the vagueness as a semantic phenomenon, but they preserve truth-functionality. Some authors defend semantics with three truth-values,

[12]For an overview of how philosophical research has struggled to cope with vagueness see, for example, Williamson (1994), Keefe and Smith (1997), Keefe (2000), Sorensen (2001), Graff Fara and Williamson (2002), Dietz and Moruzzi (2010), Ronzitti (2011) or Sorensen (2016).

like, for example, Halldén (1949), Körner (1960) or, more recently, Tye (1994) and Field (2003c). On the other hand, other philosophers have proposed semantics with infinitely many truth-values, like, for example, Zadeh (1975), Machina (1976) or Smith (2008). These proposals typically defend that some of the premises in the *Sorites* arguments are not true (how far away of truth they are depends on the theory).

Epistemicist approaches Epistemic theorists claim that vagueness is a type of ignorance. They retain full classical logic, which commits them to the existence of sharp cut-off points to the application of vague predicates; that means that, according to this view, there will be a unique precise point at which a bald man becomes not bald. Consequently, as in the case of Supervaluationism but for different reasons, the second premise in the Induction *Sorites* is false. Different versions of epistemicist approaches have been defended by Cargile (1969), Campbell (1974), Sorensen (1988, 2001), Horwich (1997, 2005b) and, in its most sophisticated stance, by Williamson (1994). This last author claims that the view can be traced back to the stoic logicians, in particular to Chrysippus (Williamson 1994, chapter 1).

Both the Liar and the *Sorites* are some of the most venerable and sturdiest paradoxes in philosophy of logic and philosophy of language. We can take the Line-drawing formulation of the *Sorites* as showing that this paradox might not be so hard as the Liar, because its conclusion (that vague predicates have sharp boundaries) would be less hard to accept that the conclusion of the Liar paradox (see Field 2008, chapter 5). However, the *soritical* paradox involves a kind of predicates that are pervasive in natural languages, which makes the vagueness paradoxes, in a sense, more dangerous than the Liar.

1.7 This Dissertation

If we consider all the schematic forms of the *Sorites* paradox presented in the previous section, it seems natural to expect a unified solution for all of them. The reason for that seems to be that they use the same characteristic of vague predicates, its tolerance, in order to achieve an unacceptable conclusion; thus, all of them point at some tension in what we are willing to accept with respect to vague predicates. This seems a good enough reason to expect a unified solution for all of them.

Similarly, most of the philosophers working on paradoxes of self-reference think that Liar-like paradoxes (like the Liar, of course, but also like the Liar Cycles paradox or Yablo's) and Curry's paradox are consequences of the

same underlying phenomena and that, consequently, they should be treated in the same way (see, for instance, among many others, Field 2008, Gupta and Belnap 1993 or Zardini 2011).

In principle, the Liar and the *Sorites* seem totally unrelated paradoxes; on the one hand, self-reference (in some or other fashion) seems to play a crucial role in the former but not in the latter and, on the other hand, truth does not seem to be a vague predicate. Nevertheless, some authors have tried to offer a common solution to both kinds of paradoxes. The aim of this dissertation is to explore the some of such proposals.

In order to do that, I will look, in chapter 2, into the nature of the notion of paradox and I will present a new characterization of that notion, more general than the traditional one. Next, in chapter 3, I will explore what should be expected, first, from a solution to a given paradox and, second, from a common solution to more than one paradox. Besides, some consequences regarding Vann McGee's account will be drawn. In chapter 4 I will look into the reasons we might have to expect a common solution to a given collection of paradoxes. Moreover, I will examine Graham Priest's proposal with a view to illustrate the discussion around the notion of a common solution to different paradoxes. Next, three approaches will be examined in some depth: Jamie Tappenden's in chapter 5, Paul Horwich's in chapter 6 and Harty Field's in chapter 7.

CHAPTER

TWO

THE NOTION OF PARADOX

Traditionally, an argument has been considered a paradox if, and only if:

(i) it is an apparently valid argument,

(ii) it has apparently true premises, and

(iii) it has an apparently false conclusion.

This view can be found, for instance, in Quine (1966), Cave (2009) or Cook (2013), among many others. As far as I can see, though, the traditional characterization of the notion of paradox is just too narrow; the conditions listed above are not necessary for being a paradox.

In this chapter I want to show that the traditional characterization does not apply to a certain kind of arguments that are problematic in essentially the same way as arguments that do fit the traditional characterization of being a paradox. Hence, the traditional characterization fails to capture whatever these two kinds of arguments have in common. In the last sections, I will offer and evaluate a more general characterization of the notion of paradox that includes both the arguments that satisfy the traditional one and those arguments that, even if not satisfying it, are problematic essentially in the same way.

2.1 A Minor Point

Before showing that the clauses in the traditional definition are not necessary, let me make one minor point regarding this definition. As it stands, the traditional definition can easily be understood in a way that does not provide

us with jointly sufficient conditions for being a paradox. Notice that, in the traditional characterization, it is used the word 'apparently' (see, for example, Priest 2006, p. 9, López de Sa and Zardini 2007, p. 246 or Soames 1999, p. 50) to convey the idea that, since a paradox seems to be a valid argument with true premises and false conclusion and that is something that is not possible, some of these have to be just apparent. There is some sense, though, in which an argument can be apparently valid which is not meant when characterizing what a paradox is. The following would be a schema of an argument that is apparently valid in some reasonable sense of 'apparently':

Argument 1

P1 If Paris is in Egypt then it is in Africa;

P2 Paris is not in Egypt.

C Hence, Paris is not in Africa.

The fact that *Denying the Consequent* is a fallacy shows that any argument following the structure above must be somewhat appealing, which means that, at least in some sense, it must be apparently valid.

We can also consider the more interesting case provided by the following argument, used by the psychologist Johnson-Laird in Johnson-Laird and Savary (1999):

Argument 2

P1 One of the following claims is true and the other is false:

- if there is a King in the hand of cards, then there is an Ace,
- if there is not a King in the hand of cards, then there is an Ace;

P2 There is a King in the hand of cards.

C Hence, there is an Ace in the hand of cards.[1]

[1]In the experiment, Johnson-Laird presented the premisses of the argument to the participants and asked what, if anything, followed from them. 100% of the participants draw the erroneous conclusion that there were an ace in the hand of cards (Johnson-Laird and Savary 1999, p. 208). I take that to show that they would have accepted the argument above as valid.

Although this argument is not valid, most of the people, as showed by Johnson-Laird and Savary (1999), considered it as valid. So this argument is another *apparently valid argument* in a relevant sense of being apparently valid. But this is not the sense intended in the characterisation of what a paradox is. It might be slightly better to use the following definition. An argument is a paradox if, and only if:

(i) it is an intuitively valid argument,

(ii) it has intuitively true premises, and

(iii) it has an intuitively false conclusion.

In the sense used here, an intuitively valid argument is an argument such that, declaring it invalid, implies giving up strong intuitions about logic. The idea is that when we accept arguments 1 and 2 above as valid we are just making a mistake and, once we realise that, we are willing to deny their validity; thus, the arguments do not challenge our basic intuitions about logical validity. That is why they do not count as paradoxes according to the last definition.

But that is not what happens when we are in front of a paradox; in these cases, even though we realize that there is something wrong, we are yet not willing to deny the validity of the arguments involved, precisely because denying it would imply giving up some of our core intuitions about validity. That is why we say that a paradox do challenge our understanding of validity.

The same happens with the other conditions of the traditional characterization. Thus, the premises of a paradox are intuitively true in the sense that their being not true would violate some of our core intuitions with respect to some of the concepts involved in them, and the same can be said about the intuitive falsehood of the conclusion. In the first case, then, if we solve the paradox by claiming that some of its premises are not true, our previous acceptance of their truth cannot be explained by a simple mistake; and analogously with respect to the falsehood of the conclusion in the second case. In other words, to change our mind with respect to the truth value of the premises (or the conclusion) we need to throw away some of the core intuitions of some of the concepts involved in the paradox.

2.2 The Traditional Definition

As we have just seen, thus, the traditional definition of the notion of paradox is the following one:

> *Definition 1.* A paradox is an intuitively valid argument with intuitively true premises and an intuitively false conclusion.

As I said, Definition 1 is too narrow. Let's see now why. I will offer one argument that is a paradox but that does not satisfy Definition 1: Curry's paradox. Suppose we have a *Curry sentence* γ, which is a sentence that asserts that if itself is true then snow is white. Then, given the T-schema, the following sentence is certainly true:

(1) $T\ulcorner\gamma\urcorner \leftrightarrow (T\ulcorner\gamma\urcorner \rightarrow \text{snow is white})$

and so is this weaker one (its left-to-right direction):

(2) $T\ulcorner\gamma\urcorner \rightarrow (T\ulcorner\gamma\urcorner \rightarrow \text{snow is white})$

Now we can apply contraction (the principle that says that from any sentence of the form '$A \rightarrow (A \rightarrow B)$)' you can infer '$A \rightarrow B$') and obtain:

(3) $T\ulcorner\gamma\urcorner \rightarrow \text{snow is white}$

Again, by the right-to-left direction of (1) we get:

(4) $(T\ulcorner\gamma\urcorner \rightarrow \text{snow is white}) \rightarrow T\ulcorner\gamma\urcorner$

And now we have almost finished. Apply *Modus Ponens* to (3) and (4) and get

(5) $T\ulcorner\gamma\urcorner$

Finally, we just have to apply *Modus Ponens* again to (3) and (5) and conclude that snow is white. Notice now that according to Definition 1, the argument above is not a paradox, for the conclusion achieved is not unacceptable. I claim that the argument just presented, though, is a paradox.

In general, the argument above could have been presented with a variable ranging over sentences in the position of 'snow is white'; then, it would not have been a paradox properly, but an schema whose instances would have been arguments. My claim is that, independently of the interpretation of the sentence variable in the Curry's sentence, all instances of the schema would have been paradoxes. But, since A could have been a true sentence, we conclude that we will have paradoxes, as the one offered above, with true and acceptable conclusions, which means that these paradoxes will not satisfy Definition 1.

Let us present, next, another counterexample to Definition 1 which also shows that it is too narrow. Consider, for example, the following argument, where Alice is 100 cm tall and William is 200 cm tall:

Argument 3

1. Alice is tall, (premise)
2. if Alice is tall, so it is someone who is 1 cm shorter, (premise)
3. someone 1 cm shorter than Alice is tall, (logic)
4. if someone who is 1 cm shorter than Alice is tall, so it is someone who is 2 cm shorter than Alice, (premise)

   ⋮

151. if someone who is 149 cm shorter than Alice is tall, so it is someone who is 150 cm shorter than Alice, (premise)
152. hence, William is tall. (logic)

According to Definition 1 this is not a paradox; for it would only be a paradox, a *Sorites* paradox, in situations where Alice was tall (so that the premises would be intuitively true) and William was not (so that the conclusion would be intuitively false). My point is that, even when Alice is not tall or William is tall, the argument is still a paradox. Hence, the conditions stated in Definition 1 are not necessary conditions.

At this point, somebody could try to defend Definition 1 replying that the arguments just presented (the Curry's argument where the Curry's sentence is build using 'snow is white' or the *Sorites* argument where Alice is not tall or William is) are not paradoxes —so that they are not valid counterexamples to Definition 1—, but something different, although closely related to paradoxes. Let's call them, say, *pathodoxes* (from *pathos* and *doxa*, a kind of ill-formed opinion). The complain might go on by saying that the traditional characterization was not intended to characterize pathodoxes, but only paradoxes.

I do not think, though, this distinction helps the proponent of Definition 1; after all, the diagnosis of the problem that a paradox poses or the problem that a pathodox poses must be the same, and the solution too. That means that the phenomena underlying both kinds of arguments are the same and that research involving both paradoxes and pathodoxes helps enlighten the notions involved in them in the same way. Even more, when we are in front of a paradox we have the feeling that there is something wrong, a feeling which constitutes what it is like to be in front of a paradox. This feeling is the same in front of any Curry's case, regardless of whether the conclusion is

acceptable or not; it's just that when the conclusion is not acceptable, this feeling might be more pressing.

In any case, if we really have to differentiate between these two notions — arguments that satisfy Definition 1 and what I've been calling *pathodoxes*—, then I claim that I am just interested in characterizing a notion that includes both. I think this discussion, thus, is just a mere linguistic conundrum. Hence, when I use the expression 'paradox' I will be thinking of this more general notion. Moreover I will call the arguments that fit the traditional characterization *traditional paradoxes*. A pathodox, then, will be an argument that shares the phenomenological character of traditional paradoxes and that does not satisfy Definition 1.

2.3 Arguments and Premises

Before continuing let me discuss a variation of the Definition 1 that, although it is eventually prey of necessity problems, it is worth examining in some detail. This discussion will offer us some insights into what kind of entity a paradox is.

Lycan (2010) proposes to understand the following characterization of the notion of paradox:

> *Definition 2.* A paradox is an inconsistent set of [sentences], each of which is very plausible. (Lycan 2010, p. 618)[2]

According to Lycan, a paradox is typically obtained by putting together the premises and the negation of the conclusion in a set, so that the result is an inconsistent set (because of the paradoxical argument) with plausible sentences as elements (Lycan 2010, p. 617).

We must be careful, though, about the way we individuate the premises of a given argument. One possibility is to understand that an argument is a non-empty set of truth-bearers (for our purposes, sentences) with a (possibly empty) subset which is the set of premises and a member which is the conclusion. Often, paradoxes like The Liar or Curry's are presented as having no premises (see, for instance, Visser 1989, p. 621). If it were the case, then, according to Definition 2, the Liar paradox and Curry's paradox using $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$ in the Curry's sentence would look exactly the same, that is just a set with the negation of the conclusion (for there are no premises):

[2]Lycan formulates his definition in terms of propositions. I understand, though, that the discussion in this chapter is independent of the nature of the truth-bearers.

$$\{\neg(Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner)\}$$

They are clearly different paradoxes, though. Hence Definition 2, if we understand premises this way, would make the Liar paradox and Curry's paradox using $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$ in the Curry's sentence indistinguishable.

Of course, what happens in these cases is that there are some meta-premises that are used in the metalanguage in which we present the paradox and that are assumed in the argument. These meta-premises might be expressed in the object language by having, for example, in the case of the Liar paradox, Robinson's arithmetic (so that we can prove the Diagonal Lemma and, hence, have the necessary surrogate of self-reference) and all the instances of the T-schema of all the sentences in the language. Thus, in this case, if **Q** is Robinson's arithmetic and **T** is the T-schema (or some way of generating it), then the Liar paradox, according to Definition 2, might look like this:

$$\{\mathbf{Q},\mathbf{T},\neg(Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner)\}$$

Notice, though, that, as before, we must be careful, for in the case just presented the Liar paradox would be again indiscernible from the Curry's paradox using $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$ in the Curry's sentence.

Hence, we need to generalize the way we construct paradoxes beyond just putting together the premises and the negation of the conclusion. That is what Lycan (2010) seems to have in mind when he represents the Liar paradox as the following inconsistent set:

$$\{\neg Tr\ulcorner\lambda\urcorner, Tr\ulcorner\lambda\urcorner\}$$

This set would be a paradox because the Liar reasoning would make very plausible both that the Liar sentence is true and that it is not. Notice that Lycan does not think that the premises usually kept implicit in the metalanguage should be put explicitly in the representation of the paradox, for it would be "very unnatural" (Lycan 2010, p. 621) and we would have "different versions depending on exactly how the ultimate premises were formulated" (Lycan 2010, p. 621). That is why, says Lycan, we should weaken Definition 2 in the following way:

> *Definition 2'.* A paradox is an inconsistent set of sentences, each of which is very plausible in its own right or has a seemingly conclusive argument for it. (Lycan 2010, p. 621)

Hence, in the case of Curry's paradox, Curry's argument must serve as a "seemingly conclusive argument" if we want that paradox to be a paradox at all. Notice now that Curry's paradox with, for instance, $2+2=5$ in the Curry's sentence, should be something close to the following (where Γ is the set of premises, whatever they might be):

$$\{\Gamma, \neg(2+2=5)\}$$

Then, though, if it is a paradox at all, it must be the case that Curry's reasoning is a seeming conclusive argument for $2+2=5$. If this is the case, though, Definition 2' becomes trivial; any inconsistent set of sentences would be a paradox. This is so because any inconsistent set of sentences would be such that all its members would have a seemingly conclusive argument; that is, its Curry's reasoning. This is a general problem that any account of the notion of paradox that defends that a paradox is a certain inconsistent set of sentences must face.[3] That is, if Curry's paradox is to be a paradox at all and it is conceived as an inconsistent set of plausible or seemingly true sentences, then the Curry's reasoning must be what gives plausibility or makes seem true at least one of the sentences. But then any inconsistent set $\{A_1, A_2, \ldots A_n\}$ will count as a paradox, for each A_i will be shown to be plausible or seemingly true because of the Curry's argument using A_i in the Curry's sentence.[4]

As a matter of fact, if we look into the examples taken above —like the Liar paradox and Curry's paradox using $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$ in the Curry's sentence—, we can see that what differentiates the Liar from Curry's paradox are neither the premises nor the conclusion, but the proof than connects them.

That is why it seems much better to suppose that when we say that a paradox is an argument, what we mean is that a paradox is a proof (which, on the other hand, is what is usually meant with 'argument'). And we can think of proofs as series of sentences such that each sentence is either accepted on non-logical grounds (and these will be the premises), or it follows from logic possibly together with some previous sentences on the series. The last sentence, then, is the conclusion.

[3] See, for instance, Rescher (2001, p. 6), Horwich (2010b, p. 226) or Schiffer (2003, p. 68).

[4] Incidentally, this line of reasoning also applies to characterizations of the notion of paradox that defend that a paradox is a single sentence or proposition (see, for instance, Sainsbury 2009, p. 1) such that we have good reasons both for accepting it and denying it. In that case, Curry's reasoning would imply that any sentence such that we have good reasons for denying it would be a paradox.

In any case, even if these problems were overcome, Lycan's Definition 2, as I said, falls prey of the same problems as Definition 1, for Lycan is supposing that the premises are acceptable and the conclusion is not (so that its negation is). Then, again, the Curry's paradox that uses 'snow is white' in the Curry's sentence might look something close to the following (where, as before, Γ is the set of premises):

$$\{\Gamma, \text{ snow is not white}\}$$

which clearly does not count as a paradox according to Lycan's proposal because, although it is an inconsistent set, not all of its members are plausible. So, as in the case of Definition 1, Lycan's Definition 2 does not offer necessary conditions for being a paradox.

2.4 The Logical Form

One possible and, at first sight, natural alternative characterization of the notion of paradox could be stated along the following lines:

> *Definition 3.* A paradox is an intuitively valid argument whose *logical form* can be used to derive an intuitively unacceptable conclusion from intuitively acceptable premises.

According to this definition, Argument 3 is a paradox even when Alice is tall, for an argument with the same logical form could be used to get a false conclusion from true premises. And the same with Curry's paradox using a true sentence to build Curry's sentence.

Definition 3, though, is too broad, for compare the following two arguments:

Argument 4

1. 2 is a natural number, (premise)
2. if a number is a natural number, so it is its successor, (premise)
3. hence, 20564 is a natural number. (logic)

Argument 5

1. 2 grains of sand do not form a heap, (premise)

2. if n grains of sand do not form a heap, neither do $n+1$ grains of sand, (premise)

3. hence, 20564 grains of sand do not form a heap. (logic)

which have the same logical form.

Now, according to Definition 3, since Argument 5 allows us to infer an intuitively unacceptable conclusion from intuitively acceptable premises and since Argument 4 and 5 share the same logical form, we should claim that Argument 1 is a paradox, which is not the case. Clearly the paradoxicality of Argument 5 does not depend solely on its logical form, but also in certain properties of the predicate 'heap'.

2.5 A First Attempt

Another way out of this situation has been proposed by López de Sa and Zardini (2007):

> *Definition 4* What really seems to be of the essence [of a paradox] is that, despite the apparent validity of the argument, the premises do not appear rationally to support the conclusion (López de Sa and Zardini 2007, p. 67)

This definition, though, is unclear in a way that could result in, again, being too broad. Consider the following argument:[5]

Zebra Argument

1. This is a zebra, (premise)

2. if this is a zebra, then it is not a cleverly disguised mule, (premise)

3. this is not a cleverly disguised mule (logic)

In some reasonable sense of *not rationally supporting*, the Zebra Argument just introduced is an intuitively valid argument (is an instance of *Modus Ponens*) such that the premises do not appear rationally to support the conclusion. This argument is a prototypical case of an argument that begs the question. It seems that the Zebra Argument begs the question because someone who does not accept the conclusion will deny the evidence that supports 1 —for instance someone who thinks, precisely, that what seems a zebra is a disguised mule.[6]

[5]Tanks to Manuel Pérez Otero for suggesting this example.

[6]The *locus classicus* for the discussion on the notion of begging the question is Jackson (1984).

My point is that if we understand the notion of *not rationally supporting* as something on the lines of *not giving the right kind of reason* —which is usually taken to be one of the features of begging the question arguments; see Sinnott-Armstrong (2012, p. 179)—, something that can be typically tested in terms of *not succeeding dialectically*, then Definition 4 is too broad. For arguments like the Zebra Argument, which are not paradoxical, will count then as cases in which the premises do not rationally support the conclusion. Even more, plain circular arguments are also arguments such that the premises do not rationally support the conclusion in the sense just stated:

Circular Argument

1. snow is white (premise)
2. snow is white (logic)

A proponent of Definition 4 could reply now that the Zebra Argument and the Circular Argument are paradoxes, in particular, they are some kind of pathodox. It should be noticed, though, that circular arguments do not share the phenomenology of paradoxes; in front of them we do not feel the uncomfortableness we feel when we are in front of a paradox. Let me elaborate that to see why circular arguments are not pathodoxes.

Take, for instance, a pathodoxical *Sorites* like the one presented in section 6.4.12, with true premises and true conclusion, and take, also, the Circular argument. Notice that both are intuitively valid arguments in the sense stated before; declaring them invalid would require giving up core intuitions of the notion of logical validity. On the other hand, in each of them, in virtue of their validity, if I believe its premises I ought to believe its conclusion. This can be stated in terms of commitment; in both the pathodoxical *Sorites* and the Circular argument, if a subject believes the premises then she is committed, in virtue of the fact that she believes the premises and the validity of the argument, to believe the conclusion. The crucial difference between the pathodoxical *Sorites* and the Circular argument is that, when in front of the former I do not want to have to believe the conclusion in virtue of the validity of the argument and the fact that I believe the premises; the commitment makes me uncomfortable. In contrast, in the case of the latter, I am willing to believe the conclusion in virtue of the validity of the argument and the fact that I believe the premises; I embrace the commitment willingly. The same happens with arguments that beg the question like the Zebra argument; if I believe the premises and I accept the argument as valid, I embrace the

commitment to the conclusion willingly. This is why we should not consider arguments like the Circular argument or the Zebra argument as paradoxes.

We can conclude, hence, that Definition 4 is too broad, for circular arguments satisfy it and, as we have just seen, circular arguments are crucially different from traditional paradoxes and pathodoxes.

2.6 The Notion of Paradox

Nevertheless, the characterization given by López de Sa and Zardini (2007) seems to follow the correct track. We may try to refine it by making more precise what do we mean when we say that, when we are in front of a paradox, the premises do not rationally support the conclusion.

We have seen that what differentiates traditional paradoxes and pathodoxes on the one hand from begging question and circular arguments on the other, is the fact that we are only willing to accept the commitment that follows from the intuitive validity of the argument in the latter case. As a matter of fact, something stronger can be said. Consider, again, a pathodoxical *Sorites* with true premises and a true conclusion; we have seen that, in this case, we are not willing to accept the commitment that stems from it. Suppose now that a given subject S who accepts the conclusion, does not accept it in virtue of the argument. If we are not willing to accept the commitment that the pathodox generates, then we will not be willing to accuse S of having done something wrong. But, in this situation, we are not only unwilling to accuse S of having done something wrong, but we believe that S would have done something wrong had she believed the conclusion in virtue of the argument. Consider now the following principle regarding commitments:

(C) If there is a commitment to accept the conclusion of an argument in virtue of its validity and the acceptance of its premises, then a subject that accepts the conclusion in virtue of the argument is not doing anything wrong.

Then, since, as we just saw, in front of a pathodoxical *Sorites* we believe that a subject who believes its conclusion in virtue of the argument is doing something wrong, (C) implies that there is no commitment implied by the pathodoxical *Sorites*. This can be generalized to any paradoxical argument and, hence, any paradoxical argument is such that when we reflect on the notion of commitment and on how a subject should behave in front of a

paradox we realize that there is no commitment to accept the conclusion in virtue of its validity and the acceptance of the premises.[7]

Compare this situation with our discussion of traditional paradoxes and Definition 1 in section 6.4.12. We said that, in front of a traditional paradox we see that, although the argument seems valid, it should not be, because it has true premises and a false conclusion. We captured this situation by stating that the argument is intuitively valid, the premises are intuitively true and the conclusion is intuitively false; in the sense that denying any of these claims would involve giving up some core intuitions of either validity or some of the key notions in the argument.

We are now in front of a similar situation. Traditional paradoxes and pathodoxes are intuitively valid arguments —and, hence they make us commit to the conclusion in virtue of accepting the premises— such that they should not be, because the commitment should not be there. Similarly, in the case of Definition 1 and traditional paradoxes, the conclusion was not true but it should have been true, in virtue of the validity of the argument and the truth of the premises. As I just said, we claimed then that the conclusion was intuitively not true. So, applying the same strategy now, we can capture the situation posed by a paradox by stating that the argument is intuitively valid but, *intuitively*, there is no commitment.

Let us try to spell this out. The idea is that if a given set of premises Γ imply a sentence δ and a subject S believes this implication, then the following is the case:

(*) if S believes Γ then S is committed, in virtue of the fact that she believes Γ, to believe δ.

Now we can present the following definition of the notion of paradox:

> *Definition 5* A paradox is an *intuitively* valid argument such that, *intuitively*, (*) fails; that is, *intuitively*, someone can believe the premises while not being committed, in virtue of believing the premises, to the conclusion.

Hence, when in front of a paradox, we are not committed, intuitively, to believe the conclusion in virtue of our belief in the premises when we do believe the premises.

We can try to make this idea more precise and present it in a more compact way with the notion of *normative requirement* as used in Broome (1999). Broome defines normative requirements as relations between propositions. I will follow Broome in writing

[7] Thanks to Dan López de Sa for raising this point.

p requires q

in order to say that p normatively requires you to q. A consequence of p requiring q is that you ought to see that if p is the case so is q. We can abbreviate this as follows:

$$O(p \rightarrow q)$$

Valid arguments are cases of normative requirements; if a given set of premises Γ imply a sentence δ we say that believing each sentence in Γ normatively requires believing δ (using 'B' for 'you believe that'):

$B\Gamma$ requires $B\delta$

and, hence,

$$O(B\Gamma \rightarrow B\delta)$$

Broome understands the conditional in the characterization of the notion of normative requirement as the material conditional (Broome 1999, p. 2). This has as a consequence that the behavior of the material conditional (and classical logic) is inherited by the notion of normative requirement, with all its virtues and, more significantly in the present case, with all its defects; hence, any argument that has some contradiction in the premises or that has a valid sentence as conclusion will constitute a normative requirement, even if the premises and the conclusion are not related in any way.

More importantly, if we rephrase Definition 5 in terms of Broom's machinery, that is,

> *Definition 5** A paradox is an intuitively valid argument such that, intuitively, does not constitute a normative requirement.

then a paradox is an argument such that, intuitively, you do not ought to see that if it is the case that you believe the premises then it is the case that you believe the conclusion. That means that, if Γ and δ are the set of premises and the conclusion of a paradox, respectively, then

$$\neg O(B\Gamma \rightarrow B\delta)$$

and, hence,

$B\Gamma$ does not require $B\delta$

But if we understand the conditional in the characterization of the notion of normative requirement as a conditional material, then any Curry's reasoning that uses a valid sentence as Curry's sentence, or even some trivial arithmetic sentence —that is, any sentence that we ought to believe— will not constitute a paradox, for it will be a normative requirement; thus, in the latter case, for instance, if δ is a valid sentence it will be the case that

$$\mathrm{O}(\mathrm{B}\Gamma \rightarrow \mathrm{B}\delta)$$

independently of which sentences are in Γ. But, having in mind this broad sense of the notion of paradox we are trying to characterize, a Curry's paradox with a valid sentence in the Curry's sentence is still a paradox, for it has the characteristic phenomenology of being in front of a paradox. One way to see this could be to imagine someone trying to convince a non believer in the Law of Excluded Middle of the validity of this law using Curry's reasoning.

That is why, if we stick to definition 5* we must understand the conditional in the characterization of the notion of normative requirement,

$$\mathrm{O}(\mathrm{B}\Gamma \rightsquigarrow \mathrm{B}\delta),$$

meaning, at least in the case of theoretical reasoning, that you ought to see that if it is the case that you believe the premises then it is the case that you believe the conclusion, *in virtue of your believing the premises*.

Therefore, the idea behind Definitions 5 and 5* is that in front of a paradox there are two confronting strong appearances that make us reconsider some of our basic intuitions of some of the concepts somehow or other involved in the paradox. On the one hand, the rules that constitute our logic lead us to consider the paradox as a valid argument. But on the other hand, when we reflect on the commitments that follow from our acceptance of the premises, we realize that we are not willing to recognize it as constituting a normative requirement.

Definitions 5 and 5*, thus, can be seen as a generalization of Definition 1. In the latter case the conclusion of the paradox was something that we were not willing to accept, typically, a contradiction. Similarly, in the former case we also have something that we are not willing to accept, namely, the fact that a certain intuitively valid argument intuitively does not constitute a normative requirement. As a matter of fact, since constituting a normative requirement is a necessary condition of being a valid argument, we are then in front of a (higher-order) contradiction.

We can not only consider Definitions 5 and 5* as generalizations of Definition 1, but we can also see that what has to be expected of a solution to

a paradox with respect to the latter definitions is the same as what has to be expected with respect to the former one. Therefore, as we will see in the next section, solving a paradox in the sense defended in this text amounts to essentially the same as solving a paradox in the traditional sense; that is, either denying some of the premises, or accepting the conclusion or, finally, denying the validity of the argument.[8]

2.7 Two Consequences

According to Definition 5 (or 5*), solving a paradox must involve at least one of the following claims.

First, in what can be called a *type 1* solution, we may show that the argument is not valid. In this case, it would be immediately explained why believing the premises does not commit you, in virtue of your belief in the premises, to believe the conclusion; or, put in another way, it does not constitute a normative requirement. Ideally, we should be able to explain why, pace the fact that the argument is not valid, it is intuitively valid, so why we should abandon the intuitions about validity that are involved in its being intuitively valid and why they are so compelling.

On the other hand, in a *type 2* solution we may defend the validity of the argument. In this case, we claim that, since the argument is valid, believing the premises commits you, in virtue of believing them, to believe the conclusion; or, more succinctly,

$$\mathrm{O}(\mathrm{B}\Gamma \rightsquigarrow \mathrm{B}\delta)$$

and hence, that their not being a normative requirement, the failure of commitment to the conclusion, must be just an illusion. This illusion would be prompted, according to type 2 solutions, by the fact that we are offered a case of an argument in which the premises are intuitively true and the conclusion is intuitively false. The situation can be described as follows. From

1. $\mathrm{OB}\Gamma$ and $\mathrm{O}\neg\mathrm{B}\delta$

you conclude

2. $\neg\mathrm{O}(\mathrm{B}\Gamma \rightsquigarrow \mathrm{B}\delta)$[9]

[8] See, for instance, Sainsbury (2009, p. 1).

[9] The principles needed in this inference are (where $\rightarrow$ is the material conditional):

What a proponent of a solution of type 2 would say is that it is this inference what explains why we think that a paradox is an intuitively valid argument such that it is not a normative requirement. Then, what type 2 solutions would show is that 1 is not the case, because either $\neg \mathrm{OB}\Gamma$ (because some of the premises are not true) or $\mathrm{OB}\delta$ (because the conclusion is, after all, true) and, thus, we do not have to conclude 2 and, consequently, we can accept the validity of the argument.

This characterization has a somewhat unexpected consequence. In order to adopt a type 2 solution you need to have been confronted to a paradox that fits Definition 1 (*i.e.* the traditional characterization) for, if not, you will not be able to identify the truth of the premises and the falsity of the conclusion as the culprits of your impression that the paradoxical argument does not constitute a normative requirement. In other words, if you are in front of a paradox for the first time, and the paradox is one of the examples we have seen like the Curry's paradox with a true sentence in the Curry's sentence, your first impression will be to blame the logic, not the truth value of the truth-bearers involved in the argument.

I think this is phenomenologically accurate. It would be very difficult to blame the inductive hypothesis in a *Sorites* argument that proceeds, say, from true premises to a true conclusion although we would still have the impression that the argument does not constitute a normative requirement.

2.8 Two Objections

In this last section I want to look at two possible problematic issues that can arise with respect to Definitions 5 and 5*.

First, let us see a reply that a proponent of Definition 1 could give in response to Curry's paradox. Recall that we showed that Curry's paradox with a true sentence in the Curry's sentence was a counterexample to Definition 1. At this point a defendant of Definition 1 could say that when we use a certain sentence A when formulating Curry's paradox the unacceptable conclusion we achieve is not A but that A follows from the premises in the Curry's reasoning.[10] We would still have, then, an intuitively valid argument with intuitively true premises and an intuitively false conclusion (namely,

$\mathrm{O}(\alpha \to \beta) \to (\mathrm{O}\alpha \to \mathrm{O}\beta)$

$\mathrm{O}\neg\alpha \to \neg\mathrm{O}\alpha$

which seem perfectly reasonable, and the fact that $O(B\Gamma \rightsquigarrow B\delta) \to O(B\Gamma \to B\delta)$.

[10]This seems to be how Cook (2013, p. 11) understands Curry's paradox.

the claim that A follows from the premises in Curry's argumentation). This reply seems to work with true sentences like 'snow is white'; that is, if we run Curry's paradox with 'snow is white' in the Curry's sentence and we conclude 'snow is white', we can read the paradox as concluding that 'snow is white' follows from the premises in the Curry's reasoning. Since this last claim seems intuitively false, Definition 1 is vindicated.

But notice that, even granting that understanding of Curry's paradox, we can also build a Curry's paradox using a logically valid sentence or a true arithmetical sentence, say, B. Then even if we understand Curry's paradox as having as conclusion that B follows from its premises, such conclusion will no longer be intuitively false, but plainly true; for, in the case of B being a valid sentence, it will follow from the premises in the Curry's reasoning (in fact, it will follow from any premises) and in the case where B is a true arithmetic sentence, since arithmetic is present in the premises to prove the Diagonal Lemma (at least in some of the ways to formulate Curry's paradox), B will follow from them.

In conclusion, even assuming that some Curry's paradoxes can be understood in a way such that they are no longer counterexamples to Definition 1, we still can, with the use of logically valid or true arithmetical sentences, devise other Curry's paradoxes that are.[11]

The second question I want to address has to do with the Preface paradox. The Preface paradox, first introduced by Makinson (1965), asks us to consider an author of an academic book who, in the preface of the book, throws a caveat to the reader about the errors that the book surely contains. At the same time, though, she is committed to each of the assertions in the book. Thus, on the one hand she believes that each assertion made in the book, say $a_1, a_2, \ldots, a_n$, is true but, at the same time, given the knowledge of her own fallibility, also believes that the conjunction of all the assertions in the book is false; that is, $a_1 \wedge a_2 \wedge \ldots \wedge a_n$ is false and, hence, $\neg(a_1 \wedge a_2 \wedge \ldots \wedge a_n)$ is true. This can be represented in the following way (using 'B' for 'the author believes that'):

(i) Ba_1, Ba_2, ..., Ba_n (that's because the author believes all her claims in the book to be true)

(ii) $B\neg(a_1 \wedge a_2 \wedge \ldots \wedge a_n)$ (that's because the author is aware of her own fallibility)

And if we accept now the principle of agglomeration,

[11] Thanks to Sven Rosenkranz for suggesting this objection.

(Agg) $(\mathrm{B}a_1 \wedge \mathrm{B}a_2) \rightarrow \mathrm{B}(a_1 \wedge a_2)$

then from (i) we can conclude $\mathrm{B}(a_1 \wedge a_2 \wedge \ldots \wedge a_n)$. Hence, the author has inconsistent beliefs; in particular, if we suppose that $\mathrm{B}\neg\phi$ implies that $\neg\mathrm{B}\phi$ we have a plain inconsistency: $\mathrm{B}(a_1 \wedge a_2 \wedge \ldots \wedge a_n) \wedge \neg\mathrm{B}(a_1 \wedge a_2 \wedge \ldots \wedge a_n)$.

Consider now, having in mind the situation described above, the following argument:

Adjunction Argument

1. $a_1, \ldots a_n$ (premises)
2. $a_1 \wedge a_2 \wedge \ldots \wedge a_n$ (logic)

which seems to be a perfectly harmless argument. According to Definition 5, though, the instance of the Adjunction argument given by the situation described in the Preface paradox will be a paradox; for, then, believing the premises does not commit us, in virtue of believing them, to believe the conclusion —that's precisely what the Preface paradox shows; you can believe all the premises while you believe the negation of the conclusion. But even if $a_1, a_2, \ldots, a_n$ are the assertions in the author's book, the resulting instance of the Adjunction Argument is not a paradox; it is just a harmless application of adjunction (the principle according to which $\phi, \psi \vdash \phi \wedge \psi$). What this means, then, is that Definition 5 is too broad; some arguments that are not paradoxes are declared as paradoxes.

Notice, though, that the fact that the logical form of an argument seems innocuous does not mean that the argument is. Consider a soritical paradox like Argument 5 in section 2.4; its paradoxical status did not depend on its logical form —which was shared by the trivial arithmetical Argument 4—, but on certain properties of the notions involved in the argument. The case is similar with respect to the instance of the Adjunction Argument where $a_1, a_2, \ldots, a_n$ are the assertions in the author's book. In this case, the argument *is a paradox*, even if its logical form can be instantiated by perfectly sound arguments. Its paradoxical status, though, stems from certain properties of the sentences in the argument, not from its logical form.[12]

[12]Thanks to Elia Zardini for suggesting this objection.

CHAPTER

THREE

SOLVING PARADOXES

Once we have stablished what a paradox is, we need to look now at what should be expected from a common solution to a given collection of paradoxes. In this chapter I will introduce a distinction between the diagnosis of a paradox and its prevention in order to illuminate what a solution both to a single paradox and to a group of paradoxes should be.

In the final part of this chapter I will examine the consequences that clarifying what a common solution should be will imply for a well known proposal to deal with vagueness and truth; Vann McGee's.

3.1 Solving One Paradox

In the previous chapter we have seen that a paradox is an intuitively valid argument such that, intuitively, does not generate the kind of commitment we would expect it to generate. This characterization was seen as a generalization of the traditional one and, besides including traditional paradoxes, also included what I called *pathodoxes*.

We have also seen some features we should expect from a solution to a given paradox. We distinguished between the following two kinds of moves towards solving a paradox, which are essentially the same as in the case of the traditional view:

Type 1 The argument is not valid.

Type 2 The argument is valid and, hence, the lack of commitment must be an illusion. Such an illusion is explained away by showing either that some of the premises are not true or that the conclusion is not false.

I claim that any solution to a given paradox must involve one of these two strategies. For convenience, in the discussion below I will distinguish between *type 2p solutions*, where it is defended that some of the premises are not true, and *type 2c solutions*, where it is defended that the conclusion is not false.

Other taxonomies for distinguishing between different kinds of solutions have been proposed. I will discuss two of them that will help illustrate and develop my own proposal. First, Cook (2013) claims that we can try to solve a paradox by rejecting some of its premises not because of their falsity, but on the grounds that some of the concepts involved in the argument are "incoherent or faulty in some other manner"(Cook 2013, p. 20). Moreover, claims Cook, the incoherence involved in the argument might explain, not only of the rejection of some of the premises, but also the rejection of the reasoning involving the faulty concept. He calls this option the *reject-the-concept strategy*. The reject-the-concept strategy is different from just rejecting some premise (Cook calls this strategy the *reject-the-premises strategy*) or rejecting the validity of the argument (the *reject-the-reasoning strategy*), says Cook, because

> we are, on [the reject-the-concept strategy], not merely saying that a particular premise is false or that a particular logical move is mistaken, but we are instead claiming that the premise or inference is somehow nonsensical or incoherent since it involves a nonsensical or incoherent concept. (Cook 2013, p. 20)

As an illustrative example, Cook considers Aristotle's view with respect to Zeno's Dichotomy paradox. The following is a reconstruction of such paradox as Aristotle presents and discusses it in his *Physics* (Aristotle 1984, 239b, 263a):

The Dichotomy

1. Moving from a point a to a point b necessarily requires performing an infinite number of tasks in a finite time. (premise)

2. It is impossible to perform an infinite number of tasks in a finite time. (premise)

3. Moving from a point a to a point b is impossible. (logic)

Now, according to Cook, merely rejecting some of the premises would consist in, for instance, claiming that the Dichotomy is just a valid proof of the claim that it is possible to perform an infinite number of tasks in a

finite time (that is, that the second premise is false). This, though, would be different, claims Cook, from Aristotle's chosen strategy, the reject-the-concept strategy:

> As a result, Aristotle rejected the first premise of Zeno's argument, but his objection to this premise is different from the sort of move involved in an application of the reject-the-premises strategy. Aristotle rejected this premise not because he thought it was false, but because he thought that Zeno's reasons for accepting this premise were based on an incoherent understanding of infinity. (Cook 2013, p. 26)

The reason why Aristotle rejected the first premise in the Dichotomy was rooted in his distinction between *actually infinite collections* and *potentially infinite collections*. The latter are sequences of objects that can always be extended although, at any step in the construction, they are always finite. According to Aristotle the notion of a potentially infinite collection was a coherent one. By contrast, an actually infinite collection contains all its infinite members at once, or at the "same time", and Aristotle claimed that such a notion was incoherent. Now, the first premise in the Dichotomy was rejected by Aristotle because it was justified by the incoherent notion of actual infinity.

Even granting Cook's characterization of Aristotle's view,[1] it is hard to see why the reject-the-concept strategy is relevantly different from the reject-the-reasoning and the reject-the-premises strategies; rejecting as incoherent one of the concepts involved in the paradox is just one of the reasons, probably among many others, that might be adduced to reject the premise or reject the validity of the argument. Thus, Cook's distinction is confusing a strategy to solve a paradox with a reason to follow a certain strategy.

[1]This reconstruction seems to be, at the very least, incomplete. Aristotle defended, as far as we know, a kind of equivocation view about Zeno's Dichotomy. Accordingly, 'infinite' would be an ambiguous expression between the notions of actual infinity and potential infinity. Then, if we resolved the ambiguity in favor of the former we would reject the first premise while accepting the second one and, if we resolved the ambiguity in favor of the latter, we would reject the second premise while accepting the first one. In Aristotle's own words:

> Therefore to the question whether it is possible to pass through an infinite number of units either of time or of distance we must reply that in a sense it is and in a sense it is not. If the units are actual, it is not possible; if they are potential, it is possible. (Aristotle 1984, 263b3)

Aristotle's view is discussed in, for example, Cajori (1915), Vlastos (1967) or Booth (1957). A general discussion of Zeno's paradoxes is offered in Salmon (2001).

As I see it, according to Cook's reconstruction, Aristotle is just offering a reject-the-premise solution to the paradox; he is proposing to reject the first premise, on the grounds that its truth can only be justified by an incoherent understanding of infinity.

It might be argued, though, that declaring a premise false or non true is not the same as declaring a premise nonsensical, and that Cook's distinction serves us to point at this. The objection might proceed by noting that a characterization like, for example, Type 2 above, would not capture a situation where the premise is rejected in virtue of involving a nonsensical concept. In order to respond to this objection, though, it is enough to read the negation in the expressions 'not valid' and 'not true' in the Type 1 and Type 2 clauses above as an exclusion negation (call it 'not_e'), so that a nonsensical sentence (argument) is not_e true (valid).

Of course, these considerations show that any solution to a given paradox will have to include an explanation of why the argument is not valid (for type 1 solutions), why some of the premises are not true (for type 2p solutions) or why the conclusion is not false (for type 2c solutions). Even more, a solution to a given paradox should be able to explain, not only what is deceiving us and why it is deceiving us, but also why it is so compelling. Thus, in the case of a type 1 solution, it has to be able to show why the argument is intuitively valid, that is, why the intuitions that rejecting the validity of the argument make us abandon are so compelling. And, similarly, with respect to type 2 solutions; they have to be able to explain away the intuitive truth (falsity) of the problematical premises (the conclusion).

In his (2003), Stephen Schiffer has proposed another taxonomy for distinguishing between different ways to solve paradoxes that is worth mentioning. He claims that paradoxes can have a *happy-face* solution or an *unhappy-face* solution, which, in turn, can be weak or strong. A happy-face solution, according to Schiffer,

> would do two things: it would identify the odd guy(s) out —that is, it would tell us that the paradox-generating propositions weren't really incompatible or else it would identify the ones that weren't true, and then it would explain away their spurious appearance so that we were never taken in by them again. (Schiffer 2003, p. 5)

The happy-face category contains the type 1 and the type 2p sorts of solutions; either we claim that the argument is not valid ("the paradox-generating propositions weren't really incompatible") or that some of the premises are not true ("it would identify the ones that weren't true"). Notice that hitherto Schiffer ignores what I am calling the type 2c kind of solution (the reject-the-conclusion strategy, according to Cook).

More interestingly, Schiffer describes another kind of solution to a paradox, the *unhappy-face solutions*:

> An unhappy-face 'solution' is simply an explanation of why the paradox can't have a happy-face solution, and this explanation will appeal to an irresolvable tension in the underived conceptual role of the concept, or concepts, generating the paradox. (Schiffer 2003, p. 6)

Besides, claims Schiffer, an unhappy-face solution can be either weak or strong:

> A weak unhappy-face solution shows that a paradox-free concept can be fashioned to do the work we expected from the paradox-generating concept, whereas a strong unhappy-face solution shows that no such paradox-free variant is possible. (Schiffer 2003, p. 6)

The characterization of unhappy-face solution prompts, as we will see, very interesting points that we need to have in mind in order to propose a solution to a paradox, but it does not seem to characterize any proper kind of solution different from the ones already considered (as he himself seems to admit when encloses the word 'solution' between quotation marks).

The notion of an unhappy-face solution can be read, as far as I can see, in two ways. First, we can understand that Schiffer is considering some meta-claim regarding the fact that no solution can be offered in front of a paradox and, hence, such meta-claim is not adding anything to any taxonomy on solutions. Second, Schiffer's unhappy-face solutions can be understood as some special cases of type 2c solutions, where we accept the conclusion and, hence, we endorse what Schiffer calls the 'irresolvable tension' in some of the concepts involved in the paradox; this would be the case, for instance, in some of the most virulent paradoxes, like the Liar, where accepting the conclusion means accepting a contradiction. In either way, Schiffer taxonomy is not adding anything new to the type 1 and type 2 classification.

Nevertheless, the distinction Schiffer draws between weak and strong unhappy-face solutions brings up an important point that cannot be ignored. When in front of a paradox, we can ask ourselves how the concepts involved in it will be changed by the solution in order to avoid the paradoxical result. These changes might involve some of the concepts explicitly present in the paradox, or some other concepts that might not appear explicitly like, for example, the notion of logical validity. As Schiffer claims, once these changes have been made we can check whether the resulting concepts can be properly considered as mere revisions of the old ones or rather they represent a deeper change in the understanding of such concepts.

These last considerations are what Charles Chihara, in his (1979), tries to capture with what he calls *the prevention of a paradox*.

At least since Tarski,[2] philosophers have thought of paradoxes as being analogous to illnesses. This analogy makes it natural to call the kinds of strategies described in the Type 1 and Type 2 clauses *the diagnoses* of a paradox (Chihara 1979, p. 590). But, as Chihara notices, this cannot be all; we also need to be able to devise, when necessary, safe paradox-free environments for the concepts involved in the paradoxes. The prevention of a paradox, thus, is the logico-semantic frame we need to adopt in order to block the paradox.

As Chihara claims, the diagnose of a paradox and its prevention are independent, in the sense that having found one of them does not imply having solved the other. As I see it, only the diagnose is a necessary condition for having a solution to a paradox; that is so because, in some cases, the prevention might not be needed. We need to issue a caveat here; I want to stress the fact that a solution to a paradox might not prompt the necessity of a prevention. For example, epistemicist solutions to the *Sorites* (like Horwich's one in chapter 6) typically endorse classical logic to deal with vague predicates —as a matter of fact, they take preserving classical logic as one of their main advantages— and do not need any change in the language or in the logic. In this sense, the prevention is not needed and, since epistemicists still are putting forward a solution, this means that having a prevention is not a necessary condition for being a (full) solution to a paradox. I could have presented this situation differently and claimed that, although offering a prevention is a necessary condition for having a solution, such a prevention might be vacuous, in the sense that the prevention consists in the logico-semantic frame we already had. I do not see any important point hanging on these two different ways of introducing diagnoses and preventions. But in connection to that it is important to notice that although in general having a prevention is not necessary in order to being able to offer a solution to a paradox, in a particular case it might be; if the diagnose of a solution implies the necessity of a prevention and the solution does not offer it, then the solution will be incomplete. I am leaving the characterizations of the diagnostic and the prevention of a paradox somewhat vague. I do not think they can be made much more precise; in any case, I expect that the examples we will see throughout this dissertation will help enlighten them.

In sum, a solution to a paradox must offer a diagnose of the paradox, which will consist of declaring the argument not (or, if you prefer, not_e) valid, in which case we will have a type 1 solution, or of declaring the argu-

[2] "The appearance of a [paradox] is for me a symptom of disease"(Tarski 1969, p. 66)

ment valid, in which case we will have a type 2 solution. In turn, when the argument is declared valid, either we can show that some of the premises are not (not_e) true or that the conclusion is not false.

Besides the diagnostic of a given paradox, a solution might have to offer a prevention of the paradox in question, which will consist of a safe paradox-free environment. The proposal of this section can be represented by figure 3.1 below.

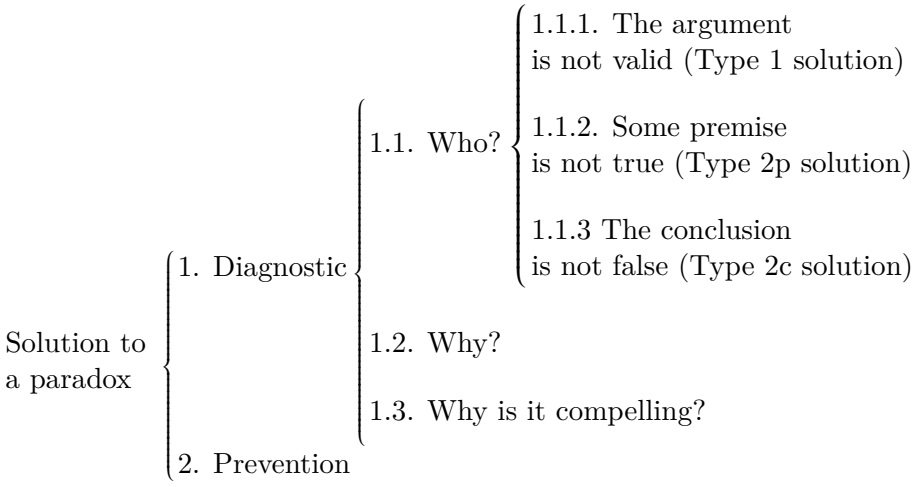

Figure 3.1

3.2 Solving more than One Paradox

Let us turn now to what features should we expect from a common solution to a given set of different paradoxes.

Think of a screwdriver; it can be used to fix, say, an electrical problem and a plumbing problem. Would we say, though, that the screwdriver is a common solution to both problems? We may say it, but it would be hardly illuminating; anyway, it is not this sense of 'common solution' that I have in mind; rather, I want to be able to see that the paradoxes can be solved by a common solution because the phenomena underlying them have something in common. And it is this something in common that the solution should point at. That seems to be part of the diagnostic problem posed by a paradox; the diagnose should give us the feature (or features) of the concepts involved in the paradox that are responsible for the deception in the paradoxical argument.

This does not mean, of course, that, given a paradox A and a paradox B, a common solution to both paradoxes must offer exactly the same diagnose. As a matter of fact, if the paradoxes are different enough, this will hardly be the case. Notice that even in the case of the *Sorites* paradox, a solution that would be of type 2p with respect to the Induction *Sorites*, would be of type 1 with respect to the Line-drawing *Sorites*.

So, which factor of the diagnose must be shared by the different paradoxes to which a given solution is supposed to be common? I think that the natural answer here is 1.2 in figure 3.1; for a solution to be a common solution to a certain group of paradoxes it must offer a common reason about why there is a deception in the paradoxes, whether in some of the premises, in the conclusion or in the validity of the argument. Any solution to a paradox will involve the necessity of rejecting some well entrenched and strong intuitions governing some of the concepts involved in the paradox. A common solution to a collection Γ of paradoxes must offer a common reason why we must reject these strong intuitions (that will very likely involve different concepts in different paradoxes) concerning the members of Γ. I do not see, in principle, that the reason why the deception is compelling should also be common, although it might be natural to expect so.

Let us see another example. Consider the following four paradoxes:[3]

1. A conditional *Sorites* with the predicate 'bald',
2. a line-drawing *Sorites* with the predicate 'bald',
3. a phenomenological *Sorites* with the predicate 'looks red',
4. the Liar.

Now, consider the following claims:

(i) 1 and 2 should have different solutions.

(ii) 1 and 3 should have different solutions.

(iii) 1 and 4 should have different solutions.

We would very likely reject claim (i) out of hand, because it seems clear that 1 and 2 should both have a common solution. With respect to claim (ii), we might not want to reject it out of hand but, still, we would be expecting some explanations why 1 and 3 do not have a common solution. Finally, we would

[3]Thanks to Dan López de Sa for suggesting this example.

probably be inclined to accept claim (iii) and we would expect further reasons to deny it. As far as I can see, the best way to explain our reactions to claims (i)-(iii) is the fact that we expect a common solution to give a common reason for paradoxicality. Thus, we think it is very unlikely that the reasons of the paradoxicality behind 1 and 3 could be different in one case and another, so that is why we would not accept claim (i). Moreover, it might be the case that there were something crucial in the phenomenological character of 3 that made plausible that the root of its paradoxicality was different from the root of 1's paradoxicality; that is why we do not reject claim (ii) immediately, although we expect some explanations. These explanations should provide us with good arguments to expect that the reason why 1 is paradoxical is different from the reason why 3 is paradoxical. Finally, we do not think, in principle, that 1 and 4 share the reason why they are paradoxes and, hence, that is why we would not expect, at first sight, a common solution to both of them. So these examples suggest that, as I said, for a solution to be a common solution to a certain group of paradoxes it must offer a common reason about why there is a deception in the paradoxes.

What about the prevention of the paradoxes? Should a common solution to a certain group of paradoxes offer the same prevention to all of them? In answer to that question I propose to distinguish between a strong common solution and a weak common solution. A *strong common solution* to a group of paradoxes will offer a unified diagnose and a unified prevention; that is, it will point to some features of some of the concepts involved in the paradoxes that will explain the source of the paradoxicality and it will offer a unified paradox-free model for such concepts. A *weak common solution* to a group of paradoxes will just offer a unified diagnose.

It is clear from what has been said that the only necessary condition to have a common solution is to have a unified diagnose. As Mark Colyvan puts it: "we want to treat the disease, not merely the symptoms" (Colyvan 2009, p. 35).

In the next section I am going to use the characterization of common solution we have discussed so far in order to argue that the approach in McGee (1991) is not a common solution to the Liar and the *Sorites*.

3.3 Van McGee: Truth as a Vague Notion

3.3 The *Sorites* in Partially Interpreted Languages

McGee (1991) uses partial predicates in order to cope with vagueness and, hence, explain away the *Sorites* paradox. Suppose you have a first order

language $\mathcal{L}$ to which you want to add a partially defined predicate P not present in $\mathcal{L}$ and suppose that $\mathcal{L}^+ = \mathcal{L} \cup \{P\}$. A *partial interpretation*, then, is an ordered pair $\langle \mathcal{M}, \Gamma \rangle$ such that $\mathcal{M}$ is a model for $\mathcal{L}$ and Γ is a first-order theory in $\mathcal{L}^+$. $\mathcal{M}$ fully interprets $\mathcal{L}$ while P is left uninterpreted. Γ is to be understood of as the theory that specifies the meaning of P.[4]

Then, according to McGee, a sentence ϕ of $\mathcal{L}^+$ is *definitely true* under the partial interpretation $\langle \mathcal{M}, \Gamma \rangle$ (in symbols $\langle \mathcal{M}, \Gamma \rangle \models \phi$) if, and only if, it is true in any expansion of $\mathcal{M}$ to a model of Γ. Accordingly, ϕ is *definitely untrue* if, and only if, $\langle \mathcal{M}, \Gamma \rangle \models \neg\phi$. An *expansion* of a first order model is obtained by adding new symbols to the language and specifying the references of these new symbols while leaving unchanged both the meaning of the old symbols and the domain of the model. In McGee's own words, $\langle \mathcal{M}, \Gamma \rangle \models \phi$ if, and only if

> ϕ will come true under any method for assigning references to the remaining symbols which makes the sentences in Γ all come out true. (McGee 1991, p. 150)

Thus, in the case we were considering, Γ will determine which interpretations of P are admissible, that is, the ones that make true all sentences in Γ.

If we apply these ideas to vague predicates, what we obtain is, essentially, Supervaluationism. As I said in page 17 Supervaluationism is a view about vagueness that was first developed in Fine (1975) using some work in van Fraassen (1966). We will see in some detail Supervaluationism in chapter 5, but let us now see the basics of it.

Suppose that the predicate we want to introduce in $\mathcal{L}$ is a vague predicate, say, 'bald'. Then, Γ would be the set of sentences governing the meaning of 'bald'. Γ would contain sentences that would determine which of the objects of the domain are clear cases of being bald and which are clear cases of not being bald; it would also contain sentences like 'if somebody with n hairs is not bald, then somebody with $n+1$ hairs is not bald either', and so on. All the possible expansions of $\mathcal{M}$ to a model of Γ, then, would be all possible admissible ways —in the sense of making true all sentences in Γ— of making completely precise 'bald'. Supervaluationists call these expansions of $\mathcal{M}$ *precisifications*. Although Fine (1975) considered the precisifications as primitive, McGee, as we have seen, considers the theory Γ governing the predicate as primitive and defines admissible precisifications in terms of it. In chapter 5 we will wee how Jamie Tappenden also considers the sentences governing the meaning of the predicate, which he calls *preanalytic*, as primitive.

[4] McGee uses the technical terminology of Chang and Keisler (1973).

The main reasons McGee offers in favor of applying Supervaluationism to vagueness are, first, that it preserves classical logic, which is seen as an advantage because "the basic logical properties of vague predicates are not appreciably different from those of precise predicates" (McGee 1991, p. 155). Second, since McGee considers vague predicates as predicates whose applicability is left indeterminated by the rules of our language, Supervaluationist semantics fits in naturally with the use of partially interpreted languages in order to capture the semantics of vague predicates (McGee 1991, p. 8).

Although McGee (1991) does not explicitly cope with the *Sorites* paradox, what we have seen up to now shows that the diagnostic for this paradox is based on the idea that vague predicates are partial predicates, in the sense that our linguistic conventions do not assign any truth value to ascriptions of vague predicates to their borderline cases. The prevention of the *soritical* paradox uses the Supervaluationist semantics, which allows McGee to preserve classical logic. For my purposes here it is not necessary to go deeper into the details.[5] So let us proceed now with the Liar paradox.

3.3 The Liar and vagueness

McGee (1989, 1991) defends an inconsistency view about the Liar. He claims that our naive understanding of truth is inconsistent and that, consequently, it has to be replaced by a new, consistent and scientifically precise notion. I will present, first, what I understand to be McGee's diagnostic of the Liar and, afterwards, I will introduce his prevention to that paradox. The main goal of this section is to show that, as it stands, McGee's proposal cannot be considered a common solution to the *Sorites* and the Liar.

The Diagnostic As McGee (1991) points out, we all tend to agree, at first sight, that any theory of truth must at least imply something on the lines of the T-schema; that is, says McGee, because, pretheoretically, we see truth ascriptions as expressing the same thought as the sentences they are ascribing truth to. Theories of truth that contain the T-schema unrestrictedly are sometimes called *naive theories of truth*. What the Liar shows is that no naive theory of truth can be right, because the former reveals how the latter is "inconsistent with manifestly observable empirical facts" (McGee 1991, p. 2); namely, the fact that $\lambda =$'λ is not true'. Thus, McGee claims:

> The logical paradoxes show that our naive understanding of truth,

[5] A thorough defense of Supervaluationism can be found in Keefe (2000). Some criticisms can be found in, for instance, Williamson (1994, chapter 5).

> which includes the acceptance of the [T-schema], is inconsistent. (McGee 1989, p. 532)

Hence, McGee's diagnostic of the Liar paradox consists in rejecting the unrestricted use of the T-schema that stems from our naive understanding of the notion of truth:

> [M]y response to the diagnostic problem is short and simple: theories that have observably false consequences are incorrect; this rule applies to informal prescientific theories no less than to scientific ones. (McGee 1991, p. 2)

McGee endorses what Charles Chihara calls *the inconsistency view of truth*, which, besides Chihara himself, has been suggested by a number of authors (see, specially, Tarski 1983, Chihara 1979, 1984, Eklund 2002 and, more recently, Scharp 2013). According to Chihara, the *consistency view of truth* is a basic position in the debate about truth which is usually taken for granted and, hence, it is hardly argued for. Such view holds that

> an accurate statement of what 'true' means will be *logically consistent with all known facts*, and in particular with all known facts of reference. (Chihara 1979, p. 607)

The inconsistency view of truth holds that the consistency view is wrong, and that truth is an inconsistent notion, in the sense that it is inconsistent with certain facts about reference. This is the position adopted by McGee as can be seen from the quotes above.

The inconsistency diagnostic is not free of problems. Specially, it has to offer a plausible explanation of how we manage to coherently use an inconsistent predicate without being led to accept unreasonable conclusions. In any case, I do not wish to investigate this issue here. For my purposes it is enough to have succeeded in showing that McGee defends an inconsistency view of truth as diagnostic to the Liar.[6]

The Prevention According to McGee, once we realize that our naive pretheoretic understanding of the notion of truth is inconsistent we must

> replace our demonstrably incorrect prescientific theory of truth with a scientific theory that is consistent with the evident empirical and mathematical facts. (McGee 1991, p. 3)

[6]It is worth noting that McGee is not usually presented, in the literature on paradoxes, as an author defending the inconsistency view (see, for instance, Eklund 2002, p. 252, Patterson 2009, p. 387 or Scharp 2013, p. 22).

The prevention of the paradox, then, will consist of a new theory of truth that, according to McGee, must satisfy the following three constraints (McGee 1991, p. 158):

I Material adequacy requirement

II Ordinary usage requirement

III Integrity of language requirement

I will sketch them in the following.

The father of the contemporary debate on the Liar paradox is Alfred Tarski. Tarski (1944, 1969, 1983) look for a definition of truth which must be formally and materially correct. The latter condition means that our definition must specify the actual meaning of the notion of truth. Tarski takes as a point of departure what he considers the most natural definition of truth:

> A true sentence is one which says that the state of affairs is so and so, and the states of affairs indeed is so and so. (Tarski 1983, p. 155)

This characterization, though, can hardly be considered appropriate, for it is not clear and precise enough. Nevertheless, it expresses the underlying idea of what a semantical definition should be and the task that Tarski has in mind is to make this idea more definite and formally correct.

It is in this framework that Tarski proposes the T-schema (he calls it *equivalence of the form (T)*) as a necessary condition for being a materially correct definition of truth. So a definition of truth for a language $\mathcal{L}$ (the object language) in a language $\mathcal{L}'$ (the metalanguage) is materially correct if it implies the following biconditional for each sentence ϕ of $\mathcal{L}$:

X is true (in the language $\mathcal{L}$) if, and only if, p

where X has to be replaced by the standard name in $\mathcal{L}'$ of ϕ and p has to be replaced by the translation in $\mathcal{L}'$ of ϕ.

What the Liar paradox shows is that, if the object language and the metalanguage are the same, and they are rich enough to achieve self-reference, then no materially adequate definition of truth can be offered. Tarski's strategy to cope with the Liar consists, essentially, in accepting the impossibility of having the language identical with the metalanguage when defining truth. In contrast, constraint III demands that McGee's truth theory for partially

interpreted languages must be given within the languages themselves.[7] This means, of course, that the conditions for material adequacy must be weakened. Let us see exactly how McGee weakens Tarski's requirement by replacing it with requirement I above, the Material adequacy requirement.

As I said, McGee's prevention for the Liar consists in introducing a new consistent and scientifically precise truth predicate. In order to do that, he proposes to treat 'true' as a vague predicate:

> I shall develop a formal model of the logic of vague terms [(partially interpreted languages)], then use this formal model to give a theory of truth which treats 'true' as a vague term. (McGee 1991, p. 7)

McGee claims that vague predicates are such that the rules governing their meaning leave indeterminate some of their ascriptions. This, he claims, is not always the case with truth, for, sometimes, the rules determining the applicability of 'true' yield conflicting answers; thus, notably in the case of the Liar sentence, the rules governing 'true' determine that such sentence is true and also that it is false.[8] McGee's proposal, hence, is to "adopt a reformed usage of 'true' which treats all the problematic cases as unsettled" (McGee 1991, p. 8).[9]

If 'true' must be seen as a vague predicate, that means that the tripartite division definitely true/definitely untrue/unsettled also applies to truth ascriptions as well as to ascriptions of vague predicates; which means, in turn, that our naive understanding of truth must be replaced by two notions: truth and definite truth. McGee goes on arguing:

> That 'true' is a vague predicate should come as no surprise. Intuitively, when we assert or deny that 'Harry is bald' is true, we are saying the same thing as when we assert or deny that Harry is bald. If that is so, then, if the linguistic conventions that govern the use of the vague term 'bald' leave it unsettled whether or not Harry is bald, the linguistic conventions that govern the use of the term 'true' likewise

[7] Most of the technical work in McGee (1991) is aimed to achieve this goal. Whether he succeeds or not is a controversial question; see Yablo (1989) and Simmons (1993, section 4.3).

[8] This is not entirely true; the Forced-March *Sorites* introduced at page 10 shows that the rules governing vague predicates can also produce conflicting answers. Thus, is we have a *soritical* series $a_1, a_2, \ldots, a_n$ for, say, the predicate 'red', we might be forced to accept, for some $1 < i < n$ that a_i is red —when we begin the forced march with a_1— and that it is not red —when we begin the forced march with a_n and we go backwards in the series.

[9] See also McGee (1989, p. 535).

> leave it unsettled whether or not 'Harry is bald' is true. [...] Thus, 'true' inherits the vagueness of all the vague nonsemantical predicates of our language. (McGee 1991, p. 217)

The sentence 'Harry is bald' is indeterminate when Harry is a borderline case of being bald, and this can be naturally explained, says McGee, with partially interpreted languages. Then, we realize that the sentence 'the sentence 'Harry is bald' is true' is also indeterminate, due to the fact that it is a truth ascription to an already indeterminate sentence —whose indeterminacy was prompted, in the first place, by vagueness. We have, hence, truth ascriptions that are already indeterminate. What McGee proposes is to treat the indeterminacy that stems from sentences like the Liar as the same as the one that indirectly stems from vagueness.

Now, if vague predicates and truth interact in the following way:

(a) If Harry is definitely bald, 'Harry is bald' is definitely true.

(b) If Harry is definitely not bald, 'Harry is bald' is definitely not true.

(c) If it is unsettled whether Harry is bald, it is unsettled whether 'Harry is bald' is true.

and 'true' must be treated as a vague predicate, then clauses (a)-(c) applied to 'true' give the following analogous clauses:

(i) If ϕ is definitely true, $Tr\ulcorner\phi\urcorner$ is definitely true.

(ii) If ϕ is definitely not true, $Tr\ulcorner\phi\urcorner$ is definitely not true.

(iii) If ϕ is unsettled, $Tr\ulcorner\phi\urcorner$ is unsettled.

Clauses (i)-(iii) constitute McGee's material adequacy conditions for truth and definite truth. McGee uses Kripke's fixed-points techniques and Supervaluational semantics[10] in order to define a truth predicate Tr and a notion of being definitely true that are materially adequate. This means that requirement I is achieved.

The second requirement, the ordinary usage requirement, is intended to guarantee that the new notion of truth agrees "with ordinary usage about the applicability of 'true' in a wide range of particular cases" (McGee 1991, p. 159). According to McGee, we use truth to convey agreement (or disagreement) to statements without having to repeat the statement again; we just

[10] We will see these techniques in some detail in chapter 6.

need to have some way of naming the statement in order to endorse it. McGee agrees with deflationists on this point and we will introduce in some detail in chapter 6 how the truth predicate allows us to express blind agreement (and disagreement). So for now an example will serve our purposes.

Suppose that Charlie is a friend of mine and that I completely trust him in political matters. Suppose further I know that yesterday Charlie gave a spirited speech on political issues. Now, although I do not know exactly what Charlie said, I want to express my full agreement to whatever he said; for I trust him blindly. One way to achieve my goal is to say:

(*) Everything Charlie said about politics yesterday is true.

In order to succeed in expressing with (*) my commitment to anything Charlie said yesterday about politics I need the following rule of inference:

(RI) $Tr^{\ulcorner}\phi^{\urcorner}$ implies ϕ

So that in committing myself to the truth of every relevant sentence ϕ I am committing, thanks to RI, to ϕ itself. RI and other similar rules of inference are definite-truth preserving given the adequacy principle stated above and, hence, they are valid in McGee's framework. (see McGee 1991, pp. 174–179 and McGee 1989, p. 534). Hence, requirement II is fulfilled.[11]

3.3 Solving the Paradoxes

It is easy to see, now, that McGee (1991) does not offer a common solution to the Liar and the *Sorites*. This is so because McGee puts forward two different diagnostics for these paradoxes; the inconsistency of our naive theory of truth for the Liar and indeterminacy based on our linguistic conventions for the *Sorites*.

We have further reasons to suppose that McGee himself would not accept the same diagnostic for the Liar and the *Sorites*. As I said, when we offer a common solution to more than one paradox we realize that their paradoxicality stems from the same root. Recall that, as I claimed at page 52, McGee thinks that vagueness always generates indeterminacy while truth can sometimes generate overdeterminacy, which suggests that the reason behind the

[11] This is also a controversial question for, some authors, like for example Field (2008, chapter 13), argue that, in order to be able to express agreement and disagreement, truth must have to obey the Intersubstitutivity Principle (see page 6), which, in the case of McGee, is not satisfied.

Liar and the *Sorites* paradoxicality might be different and, hence, the necessity of different solutions to the paradoxes generated by these concepts might be needed. What that means is that McGee might not be considering the Liar and the *Sorites* as having a common origin and, consequently, he might not see them as needing a common solution.

We conclude, hence, that, as it stands, McGee's proposal in *Truth, Vagueness and Paradox* does not offer a common solution to the Liar and the *Sorites*.[12]

[12]Notice, though, that McGee's proposal could be taken to offer a common prevention to the Liar and the *Sorites*, which, as we are going to see in the net chapter, might imply some methodological advantages.

CHAPTER

FOUR

WHY A COMMON SOLUTION

In this chapter I will also discuss when and why, in general, a common solution to more than one paradox should be expected and, in particular, when and why a common solution to the Liar and the *Sorites* should be expected. In order to help enlighten the discussion I will sketch one proposals of common solution to the Liar and the *Sorites*; Graham Priest's.

4.1 Why a Common Solution?

The following natural question is to wonder when a common solution to a given set of paradoxes is to be expected. There are at least two different groups of reasons in favor of the idea that some paradoxes might have a common solution.

First, there is one group of reasons related to methodological issues such as simplicity and uniformity. It might be argued that it is worth to seek common solutions to different paradoxes because that would be a way to deal with all of them with a minimum of resources. This is especially pressing in the case of paradoxes for, the fact that paradoxes involve very strong but incompatible intuitions implies that some high price will have to be paid in order to solve them. Hence, if we offer a single solution for some different paradoxes, we will have to pay the toll just once. In Dominic Hyde's words:

> [H]aving paid the price thought necessary to accommodate the one paradox we achieve the virtue of having to pay no additional price to accommodate the other. (Hyde (2013))

A necessary condition for achieving uniformity when dealing with several paradoxes is having a single prevention for all of them. But, as we have

seen, a common prevention is neither a sufficient nor a necessary condition for having a common solution; given some group of paradoxes Γ, we might have a single prevention for all the members of Γ without having a common solution for all of them and, conversely, we might have a (weak) common solution to the paradoxes in Γ that uses different preventative strategies for dealing with different paradoxes in Γ.[1] Nevertheless, methodological points such as uniformity and simplicity constitute a good reason for beginning the investigation, for if we achieve a strong common solution we will have gained simplicity and uniformity, and if we fail, the chances that we still get some insight to jointly prevent the members of Γ will increase.

In the second place, sometimes a certain group of paradoxes are taken to be *of the same kind* and, hence, they should be treated, it is claimed, in the same way. Graham Priest has put this idea in the form of a principle:

> *Principle of Uniform Solution, first version (PUS)*[2]
>
> If a given collection of paradoxes are of the same kind, they should all have the same kind of solution.

At first sight, PUS seems true beyond any reasonable doubt. Think, for example, of the different versions of the *Sorites* paradox offered in chapter 1; it would be very weird to defend, say, an epistemicist solution to the Conditional *Sorites* and a Supervaluational solution to, say, the No-Sharp-Boundaires *Sorites*. It would hardly make any sense because, clearly, both the Conditional and the No-Sharp-Boundaries paradoxes are of the same kind; both are *soritical* paradoxes.

But on closer inspection, PUS is not so obviously correct. The problem is that in order to assess it we need to know what counts as being the same kind of paradox and what counts as being the same kind of solution.

With respect to what counts as being the same kind of solution, there are at least two senses in which this can be understood. First, being the same

[1] Horiwch's proposal, as we will see in chapter 6, constitutes an example of this latter kind; he defends an epistemicist account both for vagueness and truth that will need a prevention only with respect to the latter. Horwich thinks that vague predicates have sharp boundaries and that, hence, the main premise of the Conditional *Sorites* is plain false. Accordingly, he claims that there is no need to change the logic and that no contradiction-free environment has to be offered. In the case of the Liar, though, Horwich needs to use something akin to the notion of groundedness and a sophisticated fixed-point construction in order to fully develop his solution.

[2] See, for example, Priest (1994, p. 32) or Priest (2002, p. 166).

kind of solution can mean just having the same prevention, in which case PUS becomes:

> *Principle of Uniform Solution, second version (PUS_P)*
>
> If a given collection of paradoxes are of the same kind, they should all have a common prevention.

Second, being the same kind of solution can be taken to mean being *the same solution*, so that if two paradoxes are of the same kind they should have a common (single) solution, in the sense stablished before, either weak or strong. In this case, which I think is the most reasonable way of understanding it, the Principle of Uniform Solution becomes:

> *Principle of Uniform Solution, third version (PUS_C)*
>
> If a given collection of paradoxes are of the same kind, they should all have a common solution.

What about what counts as being the same kind of paradox? Are the Liar and Curry's paradox the same kind of paradox? Or, thinking of the subject of this dissertation, are the Liar and the *Sorites* the same kind of paradox? These questions have no easy answer. In order to offer a fully satisfying answer to them we would need a theory of similarity of paradoxes, which is far beyond the scope of this dissertation. As a matter of fact, even a theory of identity of paradoxes is something that does not seem easily achievable at all.

Nevertheless, given a group of paradoxes, we can find some features of them that can be taken to *suggest* that they might be of the same kind or, at best, we may find some sufficient conditions for being of the same kind.

Curiously enough, some of these considerations point to a principle that is almost the converse of PUS:[3]

> *Prevention Principle of Uniform Kind of Paradox (PPUKP)*
>
> If a given collection of paradoxes have a common prevention, they are of the same kind.

PUKP is much less plausible than PUS, but it is not meant to provide knock-down reasons for establishing the sameness of different paradoxes, it is just meant to suggest that, given that some collection of paradoxes seem to

[3]See, for instance, Cook (2013, p. 194)

behave equally well in front of a given proposal of prevention, we have good reasons to think that these paradoxes might be of the same kind. PPUKP allows us to have an argument to the best explanation.

This move might seem circular or question-begging; after all, we conclude that some paradoxes are of the same kind from the fact that they have a common prevention, in order to conclude that they must have a common solution, which might imply that they have a common prevention! I do not think this is circular, though. Because, as I said, we need to understand PPUKP in a weak way, so that the fact that some paradoxes have a common prevention just suggests that they are of the same kind. We will need some other considerations in order to confidently claim that the paradoxes in question are of the same kind. Then, if we eventually accept that they are of the same kind, and we accept PUS, we cannot but accept that they must have a strong common solution.

We can be a bit subtler and, following Colyvan (2009), say that it is not the fact that they have a common prevention what makes reasonable to think of a bunch of paradoxes that they are of the same kind, but the fact that they behave similarly when tried to be solved, either by successful or unsuccessful treatments:

> *Prevention Principle of Uniform Kind of Paradox, second version (PPUKP$_P$)*
>
> If a given collection of paradoxes respond in a similar way to different preventions, they are all of the same kind.

Let us see an example, taken from Colyvan (2009), of the role this kind of considerations has in the literature. Suppose we advance a solution to the Liar that uses truth value gaps. We will see in more detail some solutions of this kind in chapters 5 and 7, but suffice to say now that they claim that the Liar sentence is neither true nor false. So we are in a situation where we have three semantic statuses for sentences: true, false and neither true nor false. As we will see in some depth in chapter 5, most if not all the solutions to the Liar suffer from what are called *revenge problems*. In this case, the revenge can be seen to come in the form of a certain sentence, usually called the *strengthened Liar*, whose construction will depend on the details of the theory used to cope with the original Liar. Even in the framework of a given theory, there might be several ways to construct such a sentence; thus, for example, when the theory posits truth value gaps like in our example, we can construct the strengthened Liar by defining a new predicate, *determinately true*, that collapses two of the semantic categories the solution is using. Hence, the true sentences will be determinately true but the false and the neither true

nor false sentences will not be determinately true. Now, using a sentence that says of itself that it is not determinately true, we get into trouble again. Therefore, in front of solutions that use truth value gaps, the Liar behaves in such a way so that new problems arise that stem from the definition of a new predicate that turns the tripartite division of the semantic statuses into a bipartite one.

The idea now is that something analogous happens with the *Sorites* paradox when the prevention used to solve it uses truth value gaps. Suppose we accommodate the borderline ascriptions of vague predicates by claiming that such ascriptions are neither true nor false. This implies, as in the case of the Liar, a tripartite division of semantic statuses; sentences can be true —the clear cases of application of the vague predicate in question—, false —the clear countercases— or neither true nor false —the borderline cases—. The problem that emerges now, and that can be seen as analogous to the strengthened Liar, is higher-order vagueness. As we saw in chapter 1 vague predicates do not only lack sharp boundaries between the clear cases and the clear countercases, but also between the clear cases and the borderline cases, and between the clear cases and the borderline cases of the borderline cases, and so on. In the first level (that is, the lack of sharp boundaries between the clear cases and the borderline cases) we can produce, given a certain vague predicate P, a strengthened paradox by introducing a new predicate, *determinately* P, so that this new predicate, as in the case of truth, collapses two of the semantic categories the solution is using: clear cases of P will be determinately P but clear counter-cases and borderline cases of being P will not be determinately P. Now, it is easy to build a new *Sorites* paradox using *determinately* P. The analogy thus established between the Liar and the *Sorites* that shows that they behave similarly when in front of a solution that uses truth value gaps can be taken as evidence to the conclusion that both paradoxes should receive the same treatment.

Another feature that has been taken, specially by Graham Priest, as a sufficient condition for being the same kind of paradox is having the same internal structure:

> *Structure Principle of Uniform Kind of Paradox (SPUKP)*
>
> If a given collection of paradoxes have the same internal structure, they are all of the same kind.

Of course, what having the same structure means or how sameness of structure can be established is under dispute. We can, for example, conclude that some collection of paradoxes have the same structure from the fact that

they share some structural properties. We find this idea applied specifically to the Liar and the *Sorites* in, for instance, the work of Jamie Tappenden. Tappenden (1993) calls attention to the fact that one can decide somewhat arbitrarily the extension of both vague predicates and the truth predicate. We will see in more detail this characteristic of vagueness and truth in chapter 5

Another important structural similarity between vagueness and truth that has been stressed in the literature is the fact that in both cases some kind or other of indeterminacy is involved. This idea can be found in many places like, for example, McGee (1991), Tappenden (1993), Field (2003c, 2008), Priest (2010a) or Hyde (2013). The indeterminacy present in both vagueness and truth can have a semantic nature (we will see examples of this in chapters 5 and **??**), an epistemic nature (we will see an example in chapter 6) or it can even be, in fact, overdeterminacy, as in the case of approaches to the paradoxes that defend the existence of truth value gluts; sentences, like that Liar and the borderline ascriptions of vague predicates, that are both true and false (we are going to see an example of such an approach in this chapter).

All these last considerations suggest that there are enough structural similarities between the Liar and the *Sorites* to at least begin the investigation into a common solution for both of them.

Apart from using structural similarities in order to support the claim that some paradoxes share the structure, we can, more directly, offer the underlying structure. This is what Graham Priest has done with all paradoxes involving self-reference. In order to further illuminate this last claim and the considerations discussed so far in this chapter, we are going to sketch in the next sections Graham Priest's proposal of a unified treatment for the Liar and the *Sorites*.

4.2 Graham Priest: Inclosures and Contradictions

4.2 The Inclosure Schema

Graham Priest has defended that many of the paradoxes in the landscape of philosophy of logic have a common underlying structure, which is taken to imply that they are of the same kind. Then, given PUS, he concludes that they should have the same kind of solution. Finally, Priest defends a solution to all these paradoxes that uses truth value gluts.

Priest has defended that paradoxes that somehow involve self-reference (the Liar among them) all have a common underlying structure that is captured by what he calls *the inclosure schema* (see, specially, Priest (1994,

2002)). More recently, Priest (2010a) has proposed that the *Sorites* also satisfies this schema, which means that both the Liar and the *Sorites* have a common underlying structure. Hence, claims Priest, they are of the same kind and, following PUS, they should be treated in the same way.

Let us present, first, the inclosure schema and, afterwards, we will show how Priest defends that the Liar and the *Sorites* satisfy it.

Inclosure Schema. There are two monadic predicates $\phi(x)$ and $\psi(x)$ and a one place function $\delta(x)$ such that:

1. There exists a set Ω such that $\Omega = \{x : \phi(x)\}$ and $\psi(\Omega)$
2. If $X \subseteq \Omega$ and $\psi(X)$, then:
   (a) $\delta(X) \notin X$;
   (b) $\delta(X) \in \Omega$.

The first thing to notice is that if there are ϕ, ψ and δ satisfying 1 and 2, then the limit case where $X = \Omega$ produces a contradiction, for then, by 2a, $\delta(\Omega) \notin \Omega$ and, by 2b, $\delta(\Omega) \in \Omega$.

Next, we need to check whether both the Liar and the *Sorites* satisfy the inclosure schema. Consider the following interpretations for ϕ, ψ and δ:

- $\phi(x)$ is the truth predicate, $Tr\ulcorner x\urcorner$,
- $\psi(x)$ is a definability predicate, so that $\psi(x)$ if, and only if, x is referred to by a non-indexical noun-phrase,
- $\delta(x)$ is the function that, given a definable set of sentences X, assigns to X a sentence σ identical to $\sigma \notin X$.[4]

With these interpretations, Ω is the set of all truths, which can be reasonably taken to be definable, so that 1 is satisfied. In order to check whether the clauses in 2 hold, suppose we have a definable set of sentences X such that $X \subseteq \Omega$. Then, $\delta(X)$ will be the following sentence σ:

(σ) $\sigma \notin X$

[4]I am simplifying and taking X as a name of itself. For readability I am also being loose on the use/mention distinction.

Now, what is essentially the Liar reasoning shows that 2a and 2b are satisfied. Notice, first, that since $X \subseteq \Omega$ and Ω is the set of true sentences, all sentences in X must be true. But if σ were in X it would not be true, for σ says, precisely, that is not in X. Hence, $\sigma \notin X$, which shows that 2a holds. But if $\sigma \notin X$, this means that $Tr\ulcorner\sigma\urcorner$ and, hence, that $\sigma \in \Omega$, which shows that 2b holds. The contradiction we get in the limit case, then, is $\delta(\Omega) \in \Omega \wedge \delta(\Omega) \notin \Omega$. Notice, besides, that $\delta(\Omega)$ is the sentence λ identical to $\lambda \notin \Omega$; now, since Ω is the set of all true sentences, λ is just claiming of itself that is not a member of the true sentences. Hence λ is just the Liar sentence and the contradiction we get from the inclosure schema is the conclusion of the Liar paradox, $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$. Hence, the Liar paradox is an inclosure paradox.

Let us turn now to the *Sorites*. In order to show that it is also an inclosure paradox, suppose $P(x)$ is a monadic vague predicate and $A = \{a_0, a_1, \ldots, a_n\}$ is a *soritical* sequence for P; that is, Pa_0, $\neg Pa_n$ and tolerance holds: for any x, $0 \leq x < n$, if Pa_x then Pa_{x+1}.

Consider now the following interpretation of the inclosure schema:

- $\phi(x)$ is the predicate $P(x)$ (restricted to A),
- $\psi(x)$ is the vacuous condition, say, $x = x$,
- $\delta(x)$ is the function that, given a set $X \subseteq \Omega$, assigns to X the first object in A that is not in X.

Interpreting the inclosure schema that way, Ω is the collection of the objects in A that are P. Notice that $\Omega \neq \emptyset$, for at least $a_0 \in \Omega$ and, besides, $\Omega \subset A$, for $a_n \notin \Omega$. Ω clearly exists and, hence, 1 is satified. Take now any set $X \subseteq \Omega$. By definition of δ, $\delta(X) \notin X$, so that 2a is also satisfied. It just remains to show that $\delta(X) \in \Omega$. We have two options, if $X = \emptyset$, then $\delta(X) = a_0$ and, hence, $\delta(X) \in \Omega$ by definition of a_0. Second, if $X \neq \emptyset$, then $\delta(X) = a_{j+1}$, where $0 \leq j < n$, such that $P(a_j)$. Then, by tolerance of P, we conclude that $a_{j+1} = \delta(X) \in \Omega$, so that 2b is also satisfied. In this case, the contradiction we obtain at the limit when $X = \Omega$ is, as before, $\delta(\Omega) \in \Omega \wedge \delta(\Omega) \notin \Omega$. Since Ω is the set of objects that are P, the contradiction we reach is that the first object in the sequence A that is not P —which is $\delta(\Omega)$— is P. So far, thus, we have seen that, according to Priest, both the Liar and the *Sorites* are inclosure paradoxes.

4.2 PUS

Priest defends that PUS is a reasonable principle (see specially Priest (1994, p. 32) and Priest (2002, p. 166)) and, furthermore, that all inclosure para-

doxes are of the same kind. Both claims imply, then, that all inclosure paradoxes should have the same kind of solution. As I said, the acceptance of PUS hangs on what we understand by *being paradoxes of the same kind*. And being a paradox of the same kind is analyzed by Priest in terms of the inclosure schema. The inclosure schema, defends Priest, captures the underlying nature or structure of the paradoxes that satisfy it and it is this sharing of the underlying structure what justifies their being of the same kind. The problem with this line of argument, though, is that when two paradoxes both satisfy the inclosure schema, what can be really concluded is that they are of the same kind *at a given level of abstraction*. I agree with Nicholas J.J. Smith that Priest does not take into account the fact that

> two objects can be of the same kind at some level of abstraction and of different kinds at another level of abstraction. (Smith 2000, p. 118)

Consider the following sentences:[5]

(i) Bill loves Ben.

(ii) Bob loves Maisy.

(iii) Nancy is standing next to Susan.

(iv) Earth orbits the Sun.

Are the facts described by them of the same type? It depends on the level of abstraction. At a lower level of abstraction, all (i)-(iv) facts are of different type; they just involve different objects. At a further level of abstraction, though, (i) and (ii) are of the same type; that is, they are facts consisting of a person loving another person. At an even further level of abstraction, (i), (ii), and (iii) are facts of the same kind; they consist of pairs of persons instantiating a relation. Finally, at a higher level of abstraction, all facts (i)-(iv) are of the same kind; they all are facts that consist of two objects instantiating a relation.

Smith proposes to reformulate PUS so that

> [p]aradoxes that share a characterization at a certain level of abstraction should indeed have solutions which likewise share a characterization at that level of abstraction. (Smith 2000, p. 119)

Hence, we can reconstruct now the principle of uniform solution following Smith's suggestion:

[5]The example is adapted from Smith (2000).

Principle of Uniform Solution, fourth version (PUS_S)

> If a given collection of paradoxes are of the same kind at some given level of abstraction, they should all have the same kind of solution at the same level of abstraction.

As Smith himself states, PUS_S is trivial. To see why, suppose we have a solution, call it C, to the Liar that somehow restricts the T-schema. At a given level of abstraction C might involve details concerning truth, truth-bearers, contexts of use, etc. The strategy C follows in order to block the Liar paradox is the restriction of the T-schema, so that the prove of clause 2 in the inclosure schema cannot go through. Hence, at a suitable level of abstraction, C will be circumventing the inclosure schema by refusing the clause 2. In principle, the level of abstraction just used to describe C, where it is considered how C blocks the inclosure schema, seems the most natural candidate to be the same level of abstraction in which the Liar and the *Sorites* satisfy the inclosure schema.

But if this is so, then any solution that solves the *Sorites* by refusing 2 in the inclosure schema will be of the same type (at the appropriate level of abstraction, which is the same at which the Liar and the *Sorites* are of the same kind) of C. But there are many solutions to the *Sorites* that solve the paradox by rejecting clause 2. We have introduced some of them in chapter 1: Supervaluationism, Epistemicism or many-valued approaches all of them solve the paradox by rejecting some of the premises in the usual forms of the *Sorites*, so that all of them block the argument that allow us to conclude, in the inclosure schema, that the first object that has not the vague property actually has the vague property. So all of them circumvent the inclosure schema by refusing the clause 2. But that means that C, Supervaluationism, Epistemicism, many-valued approaches and many other solutions to both the Liar and the *Sorites* are of the same kind in the appropriate level of abstraction in which both paradoxes are of the same type. This trivializes PUS_S.

Priest's response to the triviality of PUS_S is the following one:

> [T]he appropriate level at which to analyze a phenomenon is the level which locates underlying causes. (Priest 2000, p. 125)

According to Priest, Smith's objection to PUS was not fair in the first place, because not all the levels of abstraction are equally important; the principle of uniform solution must be applied to the level of abstraction that captures the underlying nature of the paradoxes, takes into account 'the

essence of the phenomenon' (Priest (2000, p. 124)) and 'locates the underlying causes' (Priest (2000, p. 125)) of their paradoxicality. Thus, with respect to what counts as being the same kind of solution Priest's understanding of the principle of uniform solution is closer to the one I called PUS_C, where the same kind of paradoxes should have a common solution, in the sense defended in section 3.2; for a solution to be a common solution to certain group of paradoxes it must offer a common reason about why there is a deception in the paradox. This is so because being the same kind of solution must be considered, now, at the level of abstraction that locates the underlying causes of the paradoxes. And with respect to what counts as being the same kind of paradox, two paradoxes will be of the same kind when both have a common reason about why they are deceptive; for being the same kind of paradox means now, according to Priest, to be the same kind at the level of abstraction that locates these reasons.

Putting all this together, we can see that what the principle of uniform solution should be stating is the following:

> *Principle of Uniform Solution, fifth version (PUS_P)*
>
> If a given collection of paradoxes have a common reason about why they are deceptive, they should all have a common solution.

I think PUS_P is a natural and reasonable principle, abut it is almost a platitude.[6] The fact that Priest defends PUS as a "little more than a truism" (Priest 2002, p. 287)[7] suggests that his understanding of that principle is close to PUS_P. Moreover, if Priest can show that all inclosure paradoxes share a common reason for their paradoxicality, PUS_P is sufficient to conclude that they should have a common solution; so that nothing stronger is needed. In particular, since he claims to have shown that the Liar and the *Sorites* both satisfy the inclosure schema, and he claims that the inclosure schema is the

[6] Valor Abad (2008) also claims that PUS is "trivially true" because "the expression 'same kind' is too vague" (Valor Abad 2008, p. 191). What I'm defending here, though, is that the principle is a trivial one even when we narrow down the sense of 'same kind'.

[7] Priest (2002, p. 287) compares the principle of uniform solution to an analogous principle: "same kind of illness, same kind of cure" and argues that his last principle is clearly correct:

> [I]f we have one illness in the two people, this must be due to the same cause. So the two people must be cured in the same way, namely, by attacking that cause.(Priest 2002, p. 288)

reason why they are paradoxical, he concludes that both paradoxes should have a common solution.

To my mind, the situation is the following one. In order to make PUS plausible, it must be weakened so that is becomes a platitude in the line of PUS_P. But there is a price to be paid; the notion of *same kind* must be understood then in a very strong sense: to be the same kind of paradox means to have the same underlying reason behind paradoxicality and to be the same kind of solution means to be the same (common) solution. That means that Priest needs to show that the *Sorites* and the Liar are of the same kind in this sense.

I want to argue now that the inclosure schema does not capture the reason why the *Sorites* is paradoxical. So that even granting that the inclosure schema does capture the reason why the Liar is paradoxical, which is highly debatable,[8] Priest cannot apply PUS_P to the Liar and the *Sorites*. The reason why this is so is related to the discussion in chapter 2, where the notion of pathodox was introduced. Remember that a pathodox was an argument that shared the phenomenological character of traditional paradoxes but that did not satisfy the traditional characterization, according to which a paradox is an intuitively valid argument with intuitively true premises and an intuitively false conclusion. One of the examples we saw was a *Sorites* pathodox build up with a certain series $A = \{a_0, a_1, \ldots, a_n\}$, a vague predicate P such that $\neg Pa_0$ and Pa_n and a tolerance principle stating that for any x, $0 \leq x < n$, if Pa_x then Pa_{x+1}. We concluded that the argument built in this way was essentially the same as the *soritical* argument build with a series A' like A but with Pa_0 and $\neg Pa_n$. If the two arguments thus built share the reason why they are paradoxical (in the broad sense defended in chapter 2), then, according to Priest, if one is an inclosure paradox so it should be the other. But the one built with A' is an inclosure paradox, according to Priest, and the one built with A is not.

We conclude that, at the level of abstraction in which we are trying to capture the reason why the *Sorites* is paradoxical, such a paradox is not an inclosure paradox.

4.2 Paraconsistency

In order to continue the exposition of Priest's proposal, let us suppose that he succeeded so far in showing that both the Liar and the *Sorites* are inclosure paradoxes and that this implies that they should have a common solution. Priest defends that that common solution is the adoption of a logic that allows

[8] See Grattan-Guinness (1998), Valor Abad (2008), Badici (2008) and Zhong (2012).

us to accept the contradiction $\delta(\Omega) \in \Omega \wedge \delta(\Omega) \notin \Omega$. What this means is, on the one hand, that the Liar is true and not true and, on the other hand, that the first object in the *soritical* series that has the vague property also does not have it, so that ascriptions of vague predicates to borderline cases are both true and false. A view like that that accepts that some sentences are both true and false or, alternatively, that some contradictions are true, is called *dialetheic*.

In classical logic anything follows from a contradiction; we say that classical logic has the explosion or *ex contradictione quodlibet* principle (for any sentences ϕ and ψ):

(ECQ) $\phi, \neg\phi \models \psi$

Hence, any dialetheic view will need a logic that tolerates contradictions and rejects ECQ. Such logics are called *paraconsistent logics*. Graham Priest has notoriously defended a paraconsistent approach to the Liar (see, for example, Priest (1979) or Priest (2006)) and his view has been the object of a lively and heated discussion (see, for instance, Priest, Beall, and Armour-Garb (2004) or the last chapters in Field (2008)). With respect to the *Sorites*, Priest (2010a) has defended a dialetheic framework for vagueness. For the purpose of this chapter there is no need to enter into the details and it will suffice to put forward the essentials of such a view.

Notice, first, that ECQ is an easy consequence of another logical principle also present in classical logic, *disjunctive syllogism*:

(DS) $\phi \vee \psi, \neg\phi \models \psi$

That is why paraconsistent logic also rejects DS. Now, if we define, as usual, the material conditional $\phi \rightarrow \psi$ as $\neg\phi \vee \psi$, the failure of DS implies the failure of *Modus Ponens* for $\rightarrow$. Moreover, Priest defends that the conditional used for expressing tolerance and, hence, for formulating the *Sorites* paradox, is the material one (he uses a biconditional, but nothing important hangs on this for our discussion):

> The next question is what this biconditional is. The correct understanding is, I take it, that it is a material biconditional [...]. Consecutive statements have the same truth value. This is what tolerance is all about. (Priest 2010a, p. 73)

Besides, if a conditional that obeys *modus ponens* is used in the formulation of the *Sorites*, claims Priest, "the argument [...] has less plausibility" (Priest 2010a, p. 73) because "then the truth of the sorites premises are much

less plausible" (Priest 2010a, p. 74). Priest points out that any other conditional that can be defined in a paraconsistent framework will be too strong to capture tolerance. For example, in Priest (2006, chapter 6) an entailment-expressing conditional is proposed to capture the notion of logical implication which, as such, obeys *Modus Ponens*. This conditional, though, makes the conditional premises of the *Sorites* not plausible at all; 'someone with n hairs is bald' clearly does not entail 'someone with $n+1$ hairs is bald'.[9]

To sum up, in order to be a paradox the *Sorites* must be formulated with a conditional that expresses tolerance, which, according to Priest, means that such a conditional must be the material one. That is so, he claims, because there is nothing more to tolerance than the thought that the ascriptions of the vague predicate to successive members of the *sorites* series must have the same truth value —both true or both false— and that is precisely what the material conditional expresses. Now, he goes on, since the material conditional does not obey *modus ponens* in a paraconsistent framework, he concludes that the *Sorites* argument, in its most common Conditional or Induction forms, is not valid.

Recall that, when we showed why Priest thinks the Sorites is an inclosure paradox, we interpreted $\phi(x)$ as a vague predicate P, Ω as the things in the soritical series A that are P, $\psi(x)$ as the vacuous condition and, for any $X \subseteq \Omega$, $\delta(X)$ as the first object in A that is not in X. It was straightforward to show that $\delta(X) \notin X$, for it followed directly from the definition of $\delta(x)$, but we needed a more elaborated argument in order to show that $\delta(X) \in \Omega$. In the more interesting case where we had a set $X \subseteq \Omega$ and $X \neq \emptyset$, the argument ran as follows:

Ω-*step Argument*

1. $\delta(X) = a_{j+1}, 0 \leq j < n$ (By nonemptyness of X and definition of Ω)

[9]Of course, whether there is or not a conditional obeying *Modus Ponens* and supporting *soritical* arguments is debatable. As Priest (2010a) points out, the kind of conditionals used for paraconsistent semantics and set theory (see Priest (2006, chapter 18)) are the ones used in weak relevant logics, which are such that, if they are true, then the antecedent entails the consequent. This means, as we have seen, that they are too strong to capture tolerance (Priest (2008, chapter 10)). Still, Priest is begging the question in favor of paraconsistent approaches, for, what is at stake is whether the *natural language* contains a conditional that obeys *modus ponens* and supports tolerance. If this is so, then paraconsistent logics, not only fail to offer a plausible solution to the *Sorites*, but also fail to capture the conditional in question. In any case, I will not pursue this issue here and I will suppose Priest succeeded in showing that the conditional in the *Sorites* does not obey *Modus Ponens*.

2. $a_j \in \Omega$ (By definition of A and Ω)

3. If $a_j \in \Omega$, then $a_{j+1} \in \Omega$ (a_j-instantiation of the tolerance relation)

4. $a_{j+1} \in \Omega$ (*Modus Ponens* to 2 and 3)

It can easily be seen that a conditional obeying *Modus Ponens* is used in step 3. But, according to Priest, the conditional that expresses tolerance for vague predicates does not obey *Modus Ponens* and, hence, the Ω-*step Argument* above is either not valid (if we understand the conditional in 3 as the material one and, hence, as one not obeying *Modus Ponens*) or it has one of the premises, the third one, "much less plausible" (if we understand the conditional as a stronger one obeying *Modus Ponens*). In both cases, Priest cannot use the argument in order to show that the *Sorites* paradox is an inclosure paradox.

It does not help either to claim that the conditional in 3 pertains to the meta-language, for the problem is that, in order to dialectically succeed, Ω-*step Argument* must use a a conditional that obeys *Modus Ponens* and that expresses tolerance, which, as we have seen, is the kind of conditional Priest denies to exist. Besides, if we had this conditional in the meta-language, the *Sorites* paradox would be formulated again. Since in this case, in the meta-language, we would not be able to resort to a conditional not obeying *Modus Ponens* (for then we would not be able to show that the *Sorites* is an inclosure paradox), we would have to solve this paradox using other resources, in which case we would not even have a unified solution for it.

4.2 Solving the Paradoxes

In this section we have seen that Graham Priest endorses PUS and thinks that the Liar and the *Sorites* are paradoxes of the same kind, because they have the same internal structure —captured by the inclosure schema. These considerations make him believe that the Liar and the *Sorites* need a unified treatment. We have seen, though, that PUS is a much less interesting principle than it appeared to be. Moreover, we have seen why the *Sorites* paradox cannot be counted as an inclosure paradox; first, because we have good reasons to think that the inclosure schema does not capture the internal structure of the *Sorites* and, second, because the argument Priest does to favor such claim is, by his own lights, unsound. The fact that the *Sorites* is not an inclosure paradox means that PUS cannot be applied, even in his weakened version PUS_P.

Still, we can ask ourselves whether, had it been successful, Priest's account would have constituted a common solution to the Liar and the *Sorites*.

I think the answer is yes; specifically, it would have been a strong common solution to both paradoxes. That is so because Priest offers a solution that identifies the same reason why the Liar and the *Sorites* are paradoxical; that is, their internal structure, captured by the inclosure schema. And moreover, he provides a unified prevention in the from of a paraconsistent logic.

CHAPTER

FIVE

JAMIE TAPPENDEN: TRUTH-FUNCTIONAL AND PENUMBRAL INTUITIONS

One of the first authors to entertain the possibility of a common solution to the Liar and the *Sorites* was Jamie Tappenden. In Tappenden (1993) he suggested a line of thought according to which vague predicates and the truth predicate are similar enough to support a special speech act that Tappenden called *articulation*.

In this chapter we are going to see the general framework Tappenden accepts for the Liar, which is Kripke's proposal build with the Strong Kleene semantic scheme for the logical constants. We will also present how Tappenden applies such scheme to vagueness and the *Sorites* and how, using the supervaluational framework, explains away the existence of what is usually called *the penumbral intuition*; that is, the intuition underlying the idea that there are certain sentences which we are strongly inclined to consider as true that, according to the Strong Kleene evaluations, lack truth value.

We will discuss some objections to Tappenden's approach about vagueness; some of them will be considered successful and others not. Finally, we will examine whether the ideas that stemmed from vagueness can be applied to the Liar case so that Tappenden's proposal, although not being a common solution to the Liar and the *Sorites*, might be seen as the first step towards a joint solution to them.

5.1 Kripke and *SK*

Saul Kripke presented, in his seminal paper *Outline of a Theory of Truth* (Kripke (1975)), one of the most influential approaches to the Liar paradox.

Kripke's proposal can naturally be read as proposing a logic to deal with truth that rejects the Principle of Bivalence, according to which each sentence is either true or false. Kripke presents a system where some sentences lack truth value, which, in turn, has as a consequence the rejection of the Law of Excluded Middle (LEM); that is, it rejects some sentences of the form $\phi \vee \neg\phi$.

Why, though, should we expect that rejection of excluded middle might help prevent the Liar paradox? Recall from the Introduction (page 9) that the Liar paradox allows us to conclude $Tr\ulcorner\lambda\urcorner \leftrightarrow \neg Tr\ulcorner\lambda\urcorner$ which, as we said, is equivalent in classical logic to $Tr\ulcorner\lambda\urcorner \wedge \neg Tr\ulcorner\lambda\urcorner$. Following Field (2007, p. 81), we can reconstruct how we would naturally reason in classical logic from the validity of a sentence of the form $\phi \leftrightarrow \neg\phi$ to the validity of a sentence of the form $\phi \wedge \neg\phi$:

1. $\vdash \phi \leftrightarrow \neg\phi$ (Supposition)
2. $\phi \vdash \phi \wedge \neg\phi$ (Reflexivity and *Modus Ponens* to 1)
3. $\neg\phi \vdash \phi \wedge \neg\phi$ (Reflexivity and *Modus Ponens* to 1)
4. $\phi \vee \neg\phi \vdash \phi \wedge \neg\phi$ (reasoning by cases to 2 and 3)
5. $\vdash \phi \wedge \neg\phi$ (Validity of LEM and transitivity)

This line of argumentation *suggests* that restricting LEM might help prevent the Liar paradox, although, in principle, there is no guarantee that this should be so; it is well known that intuitionistic logic, which also restricts LEM, falls prey of the inconsistencies generated by the truth predicate.

Following Hartry Field, I will call logics that restrict LEM *paracomplete logics*. In this section, we are going to present a certain version of Kripke's proposal that uses the *strong Kleene scheme* (*SK* henceforth), which does not validate LEM. In chapter 6 we will see in some detail another version that uses the supervaluational scheme (which does validate LEM and, hence, does not constitute a paracomplete approach).

The goal Kripke wants to achieve is a language with its own truth predicate; that is, he wants a language with a certain monadic predicate, Tr, that can be applied to sentences containing Tr itself and that satisfies the Intersubstitutivity Principle we introduced at page 6, which we saw should be naturally expected from a truth predicate.

In a well-known passage, Kripke states which is the intuition underlying his proposal:

> We wish to capture an intuition of somewhat the following kind. Suppose we are explaining the word 'true' to someone who does not yet understand it. We may say that we are entitled to assert (or deny) of any sentence that it is true precisely under the circumstances when we can assert (or deny) the sentence itself. Our interlocutor then can understand what it means, say, to attribute truth to (6) ('snow is white') but he will still be puzzled about attributions of truth to sentences containing the word 'true' itself. [...]
>
> Nevertheless, with more thought the notion of truth as applied even to various sentences themselves containing the word 'true' can gradually become clear. Suppose we consider the sentence,
>
> (7) Some sentence printed in the *New York Daily News*, October 7, 1971, is true.
>
> (7) is a typical example of a sentence involving the concept of truth itself. So if (7) is unclear, so still is
>
> (8) (7) is true.
>
> However, our subject, if he is willing to assert 'snow is white', will according to the rules be willing to assert '(6) is true'. But suppose that among the assertions printed in the *New York Daily News*, October 7, 1971, is (6) itself. Since our subject is willing to assert '(6) is true', and also assert '(6) is printed in the *New York Daily News*, October 7, 1971', he will deduce (7) by existential generalization. Once he is willing to assert (7), he will also be willing to assert (8). In this manner, the subject will eventually be able to attribute truth to more and more statements involving the notion of truth itself. There is no reason to suppose that *all* statements involving 'true' will become decided in this way, but most will. Indeed, our suggestion is that "grounded" sentences can be characterized as those which eventually get a truth value in this process. (Kripke 1975, p. 701)

The idea behind Kripke's proposal, hence, is that the semantical status of an ascription of truth to a sentence ϕ, $Tr\ulcorner\phi\urcorner$, will be established once the semantical status of ϕ itself is established. The process described in the quote will eventually evaluate many sentences containing 'true', but some others, like the Liar, will remain undecided. One crucial point to notice is that it must be clarified what is meant by Kripke when he states that there will be a certain set of sentences that will *eventually* get an evaluation in the process.

In order to make this claim precise Kripke proposed to consider a formal first-order language $\mathcal{L}$, the base language, as an idealization of the natural language without the truth predicate. Then, Kripke showed how that base language can be expanded with a truth predicate, Tr. In order to do that, let us suppose that we expand $\mathcal{L}$ to $\mathcal{L}^+ = \mathcal{L} \cup \{Tr\}$, a language with a new monadic predicate Tr. Suppose, furthermore, that for every formula $\phi \in \mathcal{L}^+$ we can express its canonical name $\ulcorner\phi\urcorner$ in $\mathcal{L}$ via some codification. I will suppose that $\mathcal{L}$ is strong enough to prove the Diagonal Lemma.

Given a classical model $\mathcal{N}$ for the base language with domain D, I will use $\langle \mathcal{N}, A\rangle$ to refer to the model of the expanded language $\mathcal{L}^+$ whose interpretation of Tr is A, which will be a set of codes of formulas of $\mathcal{L}^+$. I will use $|\alpha|_{\mathcal{M}} = 1$ to mean that the formula α has semantic value 1 in the model $\mathcal{M}$ (and the same for having semantic value 0).

As I mentioned before we are going to use SK, which is a three-valued scheme, $|\,|^{sk}$, that will take as semantic values 0, ½ and 1. For any $\phi, \psi \in \mathcal{L}^+$, any classical model $\mathcal{N}$ for the base language $\mathcal{L}$ and any set of (codes of) sentences X, $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle}$ is defined in the following way:

1. If ϕ is an atomic formula of $\mathcal{L}$, then $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = |\phi|_{\langle \mathcal{N}\rangle}$.
2. If $\phi = Tr\ulcorner\psi\urcorner$, then
   (i) $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = 1$ if, and only if, $\psi \in X$;
   (ii) $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = 0$ if, and only if, $\neg\psi \in X$ or ψ is not a sentence;
   (iii) $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = 1/2$ otherwise.
3. $|\neg\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = 1 - |\phi|^{sk}_{\langle \mathcal{N}, X\rangle}$
4. $|\phi \vee \psi|^{sk}_{\langle \mathcal{N}, X\rangle} = \max\{|\phi|^{sk}_{\langle \mathcal{N}, X\rangle}, |\psi|^{sk}_{\langle \mathcal{N}, X\rangle}\}$
5. $|\exists x\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = \max\{|\phi(d/x)|^{sk}_{\langle \mathcal{N}, X\rangle} : d \in D\}$, where $\phi(d/x)$ is the result of replacing all free occurrences of x in ϕ with d.[1]

The other logical constants ($\wedge, \rightarrow, \forall$) are defined in the usual way.

With SK at hand, we can now define a series of sets of sentences that represent the provisional extensions of the truth predicate that the learning subject of the previous quote is trying to grasp.

[1] I am supposing that every member d of the domain serves as a name of itself.

$$K_0 = \emptyset$$
$$K_{\sigma+1} = \{\phi \in \mathcal{L}^+ : |\phi|^{sk}_{\langle \boldsymbol{N}, K_\sigma \rangle} = 1\}$$
$$K_\lambda = \bigcup_{\alpha<\lambda} K_\alpha$$

where λ is a limit ordinal.

We will see in more detail such kind of constructions in Chapter 6, but let us say, for the moment, that Kripke (1975) showed that this construction is monotonic, that is, for any ordinals θ and ρ, if $\theta \leq \rho$, then $K_\theta \subseteq K_\rho$. As we will see, this means that there exists a fixed point of the construction; that is, there is an ordinal ρ such that $K_\rho = K_{\rho+1}$. I will call this fixed point $\boldsymbol{K}$. As a fixed point, the most important feature of $\boldsymbol{K}$ is that, under SK, when we interpret $\boldsymbol{K}$ as the extension of the truth predicate, the semantic status of a given sentence ϕ is always identical to the semantic status of its truth ascription; that is, $|Tr^\ulcorner\phi^\urcorner|^{sk}_{\langle \boldsymbol{N}, \boldsymbol{K} \rangle} = |\phi|^{sk}_{\langle \boldsymbol{N}, \boldsymbol{K} \rangle}$. This guarantees that the Inter-substitutivity Principle introduced at page 6 holds and, hence, Tr has one of the main characteristics we expect from a truth predicate. Consequently, we achieved Kripke's goal: we just presented a language with its own truth predicate that does not generate contradictions.

Returning now to Kripke's quote, notice that the zero stage in the construction, K_0, which is the empty set, captures the situation Kripke's subject is in when he does not understand 'true' at all; that is, the extension of the truth predicate is empty. Then, the first stage, K_1, contains everything having semantic value 1, provided that the extension of the truth predicate is empty; that is, K_1 is an elegant way of capturing the first step Kripke's subject has to follow in order to comprehend the word 'true', which consists of incorporating to the extension of the truth predicate everything he is 'entitled to assert' given what he knows about truth, which is nothing. And, of course, idealizing the subject to have access to all non-semantic facts, what the subject is entitled to assert at the first stage is just all the sentences that will have semantic value 1 independently of the extension of the truth predicate; that is, K_1. When we eventually reach a limit ordinal, we take stock and we just collect everything from the previous stages, so that the construction can be seen as idealizing the process of generalization at limit ordinals. The process, hence, goes on indefinitely.

Another important point in Kripke's proposal is the use of the notion of *groundedness*, which Kripke attributes to Herzberger (1970). In the previous quote, Kripke presents what can be called, following Kremer (1988), an *upwards* view of the notion of groundedness; the idea is that, from a given

collection of non-semantic facts, we can begin a process that allows us to conclude the semantic value of more and more sentences involving the notion of truth. In the quote above, Kripke is describing how such a stage-by-stage process might be accomplished. At the end, we can characterize as grounded the sentences that 'get a truth value in the process'.

Additionally, the notion of groundedness can also be explained from a *downwards* point of view. In Kripke's own words:

> In general, if a sentence [...] asserts that (all, some, most, etc.) of the sentences of a certain class C are true, its truth value can be ascertained if the truth values of the sentences in the class C are ascertained. If some of these sentences involve the notion of truth, their truth value must in turn be ascertained by looking *other* sentences, and so on. If ultimately this process terminates in sentences not mentioning the concept of truth, so that the truth value of the original statement can be ascertained, we call the original statement *grounded*; otherwise, ungrounded.(Kripke 1975, p. 693)

That is, a sentence is grounded when its semantic value eventually depends, even if indirectly, on non-semantic facts.

Kripke defines grounded sentences to be those such that they or their negation are in $\boldsymbol{K}$, which allows him to capture the intuitions embodied in both views about groundedness stated above.

It is worth mentioning here that $\boldsymbol{K}$ is not the only fixed point that can be constructed over a given base model. As a matter of fact, $\boldsymbol{K}$ is the *minimal fixed point*; there are many different fixed points that can be achieved and $\boldsymbol{K}$ is the smallest one, in the sense that it is included in all the other ones. One way to see this is to consider generalizations of the construction above that start, not from the empty set, but from other appropriate sets of sentences.[2] Consider, for example, the *Truth-teller*, a sentence τ that is identical to its own truth ascription:

(τ) $Tr^{\ulcorner}\tau^{\urcorner}$

Notice that we could define a series of sets of sentences like the one above but beginning with $K'_0 = \{\tau\}$ or also with $K''_0 = \{\neg\tau\}$ instead of K_0. In both cases we would get two different fixed points, $\boldsymbol{K'}$ and $\boldsymbol{K''}$ that would be supersets of $\boldsymbol{K}$. Although Kripke does not commit himself to any particular

[2]More specifically, the construction can start with any set of sentences X such that $X \subseteq \{\phi : |\phi|^{sk}_{\langle \boldsymbol{N}, X\rangle} = 1\}$. This is necessary to guarantee that the 0-stage of the construction is a subset of the 1-stage.

fixed point, the minimal fixed point $\boldsymbol{K}$ is usually considered the most natural candidate to be the extension of the truth predicate.[3]

Kripke's proposal with *SK*, though, is not free of problems; let me present two of them.

5.1 Weakness of the Logic

First, the use of *SK* makes the theory too weak, as *SK* fails to validate some elementary laws such as $\phi \rightarrow \phi$.[4] Actually, it is well-known that *SK* has no tautologies at all. This also means that there will be instances of the T-schema (for example, the λ-instance) that will be assigned the semantic value ½ and, hence, the T-schema will fail as a principle governing truth. A natural question we might raise now is whether a new biconditional $\Leftrightarrow$ can be added to the language so that the T-schema gets value 1; since *Tr* satisfies the Intersubstitutivity Principle, it would be sufficient to have an appropriate conditional (in a sense to be specified shortly) that validates $\phi \rightarrow \phi$, for any ϕ. Let us show now that this is not possible if we suppose some basic features that this biconditional should satisfy. If we want $\Leftrightarrow$ to be a natural generalization of the classical biconditional and we want the T-schema to have value 1 in Kripke's framework, we should expect $\Leftrightarrow$ to satisfy the following conditions, for any sentences ϕ, ψ of $\mathcal{L}^+$ and any set of sentences X:

(i) $|\phi \Leftrightarrow \psi|^{sk}_{\langle \mathcal{N}, X\rangle} = 1$ if, and only if, $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle} = |\psi|^{sk}_{\langle \mathcal{N}, X\rangle}$.

(ii) $\Leftrightarrow$ must be normal, that is, it must agree with the classical biconditional on the classical values.

(iii) $\Leftrightarrow$ must be commutative.

If we take into account these three conditions we obtain the connectives satisfying the following table:

$\Leftrightarrow$	0	1	1/2
0	1	0	a
1	0	1	b
1/2	a	b	1

[3] See, for instance, Soames (1999) or Kremer (1988).

[4] This last sentence is, as a matter of fact, just an instance of LEM, since $\phi \rightarrow \psi$ is defined as $\neg\phi \vee \psi$.

where $a, b \in \{0, 1/2\}$.[5]

Let us see now why any biconditional satisfying conditions(i)-(iii) cannot be added to $\mathcal{L}^+$ in Kripke's system. Suppose we add the biconditional $\Leftrightarrow$ to $\mathcal{L}^+$ and that we get a fixed point $\boldsymbol{X}$ with SK together with the rules for $\Leftrightarrow$, so that for any sentence $\phi \in \mathcal{L}^+$, $|\phi|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = |Tr^{\ulcorner}\phi^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle}$, where sk' is the new evaluation including $\Leftrightarrow$. Consider now the following two sentences, where ψ is any sentence such that $|\psi|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 0$:

(α) $\psi \Leftrightarrow Tr^{\ulcorner}\alpha^{\urcorner}$

(β) $Tr^{\ulcorner}\alpha^{\urcorner} \Leftrightarrow Tr^{\ulcorner}\beta^{\urcorner}$

We have now three possible cases, all leading to a contradiction.

- First, if $|Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 0$, then $|\psi \Leftrightarrow Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1$, that is, $|\alpha|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1$ and, since we are supposing that $\boldsymbol{X}$ is a fixed point, $|Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1$. Contradiction.
- Second, if $|Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1$, then, by $\boldsymbol{X}$ being a fixed point, $|\alpha|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1$ and, hence, by definition of α, $|Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 0$. Contradiction.
- Finally, suppose that $|Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1/2$, in which case $|\alpha|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1/2$. We have to consider, now, $|Tr^{\ulcorner}\beta^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle}$.
  - If $|Tr^{\ulcorner}\beta^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1$ then $|\beta|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = b$ and, hence, $|Tr^{\ulcorner}\beta^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = b$. Contradiction.
  - If $|Tr^{\ulcorner}\beta^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 0$ then $|\beta|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = a$ and, by he fact that $\boldsymbol{X}$ is a fixed point, $a = 0$. But, if $|Tr^{\ulcorner}\alpha^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1/2$ and $|\psi|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 0$, then $|\alpha|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = a$ which, by supposition, is ½. Contradiction.
  - If $|Tr^{\ulcorner}\beta^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1/2$, then $|\beta|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle} = 1 = |Tr^{\ulcorner}\beta^{\urcorner}|^{sk'}_{\langle \mathcal{N}, \boldsymbol{X}\rangle}$. Contradiction.

This means that we cannot add a suitable biconditional to the language —using Kripke's construction as it stands— in order to have the T-schema.

[5] There are two relevant connectives that satisfy this table; in the first, $a = b = 0$, which makes it strengthen condition (i) to (i'):

(i') $|\phi \Leftrightarrow \psi|^{sk}_{\langle \mathcal{N}, X\rangle} = 0$ if, and only if, $|\phi|^{sk}_{\langle \mathcal{N}, X\rangle} \neq |\psi|^{sk}_{\langle \mathcal{N}, X\rangle}$.

The other relevant connective is Łukasiewicz 3-valued biconditional, which assigns ½ both to a and b. We will discuss again Łukasiewicz logics in chapter 7.

5.1 Revenge

The second problem of Kripke's approach with *SK* I will introduce falls under what is usually called *revenge phenomena*. Most if not all approaches to the Liar paradox suffer from expressive weakness; there are some semantic notions that cannot be expressed in the language of the theory under pain of inconsistency. Kripke's proposal is not an exception. Eventually, the solution Kripke is advancing for the Liar conundrum consists in claiming that the Liar lacks truth value and hence, *a fortiori*, that it is not true. That makes natural to expect one to be able to express Kripke's solution within $\mathcal{L}^+$; that is, to express that the Liar sentence is not true, which would be asserted by $\neg Tr\ulcorner\lambda\urcorner$. Unfortunately, this is not possible, for, given a base model $\mathcal{N}$, $|\lambda|^{sk}_{\langle\mathcal{N},\mathcal{K}\rangle} = 1/2$ and, consequently, $|Tr\ulcorner\lambda\urcorner|^{sk}_{\langle\mathcal{N},\mathcal{K}\rangle} = 1/2$, which means, of course, $|\neg Tr\ulcorner\lambda\urcorner|^{sk}_{\langle\mathcal{N},\mathcal{K}\rangle} = 1/2$, so that this last sentence is not assertable. In order to be able to express that the Liar is not true, we could introduce to the language an exclusion negation, $\neg_e$, which assigns 1 to sentences with semantic value ½ (and behaves as expected for a negation with respect to the other values), so that $|\neg_e Tr\ulcorner\lambda\urcorner|^{sk}_{\langle\mathcal{N},\mathcal{K}\rangle} = 1$. Unfortunately, if we added $\neg_e$ to the language we would be able to create a new paradox (a *revenge paradox*); notice that, in this case, $\phi \vee \neg_e\phi$ would be valid and, hence, the Diagonal Lemma would allow us to create a sentence λ_e such that $\lambda_e \leftrightarrow \neg_e Tr\ulcorner\lambda_e\urcorner$. Then, by the usual Liar reasoning together with $Tr\ulcorner\lambda_e\urcorner \vee \neg_e Tr\ulcorner\lambda_e\urcorner$ we would get $Tr\ulcorner\lambda_e\urcorner \wedge \neg_e Tr\ulcorner\lambda_e\urcorner$, which is a contradiction.

Alternatively, we could try to express that the Liar is not true using a determinately operator D and ordinary *SK* negation. In order to do that, it would be sufficient that D satisfied, for any sentence ϕ of $\mathcal{L}^+$ and any model $\mathcal{M}$ for $\mathcal{L}^+$, the following conditions:

(i) $|D\phi|^{sk}_{\langle\mathcal{M}\rangle} = 1$ if, and only if, $|\phi|_{\langle\mathcal{M}\rangle} = 1$

(ii) $|D\phi|^{sk}_{\langle\mathcal{M}\rangle} = 0$ if, and only if, $|\phi|_{\langle\mathcal{M}\rangle} \neq 1$

With the use of the D and $\neg$ we could express, now, Kripke's solution to the Liar by claiming that the Liar is not determinately true, that is, $\neg DTr\ulcorner\lambda\urcorner$. But, again, D would be a source of inconsistency. Notice that, as before with $\neg_e$, LEM holds of D: for any ϕ, $D\phi \vee \neg D\phi$. Again, a sentence λ_d could be defined such that $\lambda_d \leftrightarrow \neg DTr\ulcorner\lambda_d\urcorner$, which, together with the λ_d-instance of the T-schema would yield $Tr\ulcorner\lambda_d\urcorner \leftrightarrow \neg DTr\ulcorner\lambda_d\urcorner$, which, as can be easily shown, is inconsistent with (i)-(ii) rules governing D.[6] This means

[6] As a matter of fact, we could define the determinately operator in terms of exclusion negation, so that $D\phi =_{def} \neg\neg_e\phi$.

that, under pain of inconsistency, we cannot express exclusion negation nor a determinately operator in Kripke's approach.

Tappenden (1993) proposed to apply *SK* both to truth and vagueness, so that both the Liar and the *Sorites* paradoxes could be solved. Tappenden accepted the general framework proposed by Kripke to cope with the Liar. We are going to see how Tappenden dealt with the first of the problems that affect *SK* stated above; he used a new speech act called *articulation* to explain away the apparent truth of some sentences that are assigned ½ in *SK*. In the next section we are going to see how Tappenden's approach applies to vagueness and the *Sorites* paradox. We will introduce it, following Tappenden, as a way to resolve the tension that arises between two intuitions underlying certain assignments of semantic values to sentences involving vague predicates: the truth-functional and penumbral intuitions.

5.2 Intuitions and Sharpenings

Imagine that you do not know what to say in front of the sentences 'John is tall' and 'Joe is tall' due to the fact that John and Joe are two borderline cases of being tall. You can say, then, that these sentences are gappy[7];neither true nor false. Imagine now that you are confronted to the sentence 'if John is tall, then Joe is'. As far as you know, and having in mind the truth value of its constituents, you would be unable to assign any truth value to this second sentence. Hence, it would also be neither true nor false.

Now imagine that you know that Joe is taller than John. Then, it seems that you would say that the previous sentence, 'if John is tall, then Joe is', should be true. So, which is its truth value?

There are two intuitions underlying cases like these. The first one is the truth-functional intuition: we tend to see sentential connectives as truth functions; the truth value of a sentence with a sentential connective should depend on the truth value of its parts and this value should be uniform in the sense that, if sentences ϕ and ψ have the same form and their sentential constituents have the same truth values, then ϕ and ψ should share the same truth value.

The second intuition is the penumbral intuition: there are sentences that seem almost analytic to us and we are strongly inclined to assign truth to

[7]In other chapters I will use the term 'indeterminate' in cases like these, but since Tappenden uses this term, as we will see, in a non-standard way I prefer to use, in this chapter, the expression 'gappy' to refer to sentences that are neither true nor false. Moreover, since the term 'indefinite' is frequently used interchangeably with 'indeterminate', the use of the former would also be inappropriate.

them, although, according to the truth-functional intuition, they should not have any truth value. Tappenden (1993), following Fine (1975), calls *penumbral sentences* sentences like 'if Joe is taller than John, then, if John is tall, Joe is'. In order to see how Tappenden defines the concept of penumbral sentences, we need to see first what is, according to him, a pre-analytic sentence.

One of the features of vague predicates is that their extensions can vary according to circumstances: we can increase in precision a vague predicate if it is necessary in a specific context. These increases in precision are, on the one hand, arbitrary for, usually, when a certain context demands sharper boundaries, we can choose them among a certain set of possibilities. But, on the other hand, not all increases in precision are equally valid. Tappenden proposes one example that will serve as an illustration of that. Suppose we introduce into English the predicate 'tung' whose use is governed only by these rules:

(i) 'tung' applies to anything of mass greater than 200 Kg.

(ii) 'tung' does not apply to anything of mass less than 100 Kg.

If we compare this predicate with 'heavy' we can see, first, that both behave in certain respects in the same way but, second, that they differ in a crucial one; the idea is that, provided that all our understanding of 'tung' is given by (i) and (ii), we can increase its precision in such a way that, given two objects a and b, b heavier than a and both unsettled with respect to the predicate, b counts as non tung while a counts as tung. We cannot increase the precision of 'heavy' in this way; if b is heavier than a, then our understanding of the predicate imply that, if a is heavy, so is b. We can say, then, that a precisification (a way of precisifying a predicate) is admissible if the sharper boundaries drawn are acceptable according to the meaning of the predicate.

Now, taking into account that constraints on increases in precisions can be seen as assignments of truth values to sentences, the latter example suggests that one of the collections of constraints in precision is one whose members can be formulated thus:

> Never make words w_1, ..., w_n more precise in such a way that sentence ϕ become false.

Sentences like ϕ in the example are called *pre-analytic* by Tappenden. Hence, if a sentence is pre-analytic then anyone who understands ϕ knows not to draw more precise boundaries to any expressions in ϕ in such a way

that ϕ would be false. An example of a pre-analytic sentence in Tappenden's sense is 'if Joe is taller than John, then, if John is tall, Joe is' or, following him and simplifying, 'if John is tall, Joe is'. Notice that pre-analytic sentences are never false but, depending on the semantic frame, they do not need to be always true. Moreover, Tappenden considers the notion of pre-analytic sentence as basic and the notion of admissible precisification as derived.[8] It needs to be noticed that, according to that, the characterization of pre-analytic sentences given above cannot be a definition, for pre-analytic sentences cannot be defined if they are primitive; we might say only things like that they receive its semantic status (whatever it be) as a consequence of the very meaning of the predicates or some other charaterization close to that.

Now, what Tappenden calls *penumbral sentences* are pre-analytic sentences that, relative to some assignment respecting the truth-functional intuitions, are considered neither true nor false, even though we are strongly inclined to regard them as true[9]. This tendency to consider penumbral sentences true is what underlies the penumbral intuition.

5.3 Gaps and Supervaluations

We will present, now, Tappenden's gap theory. We will do that in contrast with Supervaluationism; as we will see, the former tries to capture the truth-functional intuition while the latter elaborates the penumbral one.

Tappenden uses a partial model, called the *pre-assignment*, that assigns to any predicate P an extension, that is, a set of objects to which the predicate clearly applies, and an anti-extension, that is, a set of objects to which the predicate clearly fails to apply.

Tappenden defines having semantic value 1 (for our present purposes we can consider that having semantic value 1 is just being true) in the pre-assignment using SK. The crucial difference with the previous section is that, in Kripke's solution to the Liar the truth predicate was the only partially defined predicate, while in Tappenden's view all vague predicates are partially

[8] Cf. chapter 3, section 3.3.

[9] According to Kit Fine, who introduced the expression 'penumbral connection', a penumbral connection is a logical relation that holds among sentences that do not receive any classical truth value. Then, the truths that arise from penumbral connections are penumbral truths. According to Tappenden, though, penumbral sentences are a subset of pre-analytic sentences and, since by definition are assigned no truth value, there is no need to distinguish between penumbral sentences and penumbral truths.

defined. A sentence, then, is true if it is true in the pre-assignment and false if it is false in the pre-assignment. It can be seen now that in an account of vagueness such as this one, the truth-functional intuition plays a central role.

Supervaluationism bases truth valuation, not upon the pre-assignment, but upon the set of admissible ways of precisifying vague predicates[10] in the pre-assignment. Such precisifications must be admissible and complete. A precisification is complete when it behaves classically; that is, when there are no unsettled cases of the predicates. Additionally, a precisification is admissible when it does not make any member of the set of pre-analytic sentences false (so, in particular it must not make any member of the set of penumbral sentences false); that is, it does not conflict with our intuitions about the meaning of the predicates. The supervaluationist claims, then, that a sentence is true if, and only if, it is true on all complete admissible precisifications and it is false if, and only if, it is false in all complete admissible precisifications.

We can see now that penumbral sentences are true under this framework, for they are true on all complete admissible precisifications. Thus, in a supervaluationist account of vagueness the penumbral intuition is fully respected. Actually, Kit Fine presents it as one of the main motivations for his account; if we have a blob that is a borderline case of being red and of being pink, we must be capable of explaining why we tend to say that the sentence 'the blob is both pink and red' is false while we tend to say, when the blob is also a borderline case of being small, that the sentence 'the blob is both pink and small' is neither true nor false. It seems that no truth functional approach can explain that. But one possible way of explaining it is to take into account that if we make the relevant predicates more precise we will not be able to make the blob a clear case of 'pink' and 'red' at the same time, while we will be able to make the blob a clear case of both predicates 'pink' and 'small'. Fine dialectically presents Supervaluationism as a way of making rigorous the last suggestion that differences in truth-values reflect differences in how the predicates can be made more precise.

Now, returning to Tappenden and in order to finish the presentation of his approach, it is important to notice that, as a matter of fact, it can be seen as a position between the truth-functional and the penumbral intuitions. Let's see why.

[10]Kit Fine considers the notion of admissibility as primitive, but Tappenden, as we have seen, considers the notion of pre-analytic sentence as basic and the notion of admissible precisification as derived; thus, an admissible precisification is one that does not make any penumbral sentence false.

First, it has to be said that Tappenden uses the supervaluationist machinery we have just seen in order to solve the fact that his framework cannot distinguish between predicates like 'tung' and 'heavy'; that is, it cannot express how constraints on increases in precision are embodied in the meaning of predicates. After all, if a and b are borderline cases of the predicates 'tung' and 'heavy', and both predicates have the same extension and anti-extension, the sentence 'if a is tung so is b' has the same status as the sentence 'if a is heavy so is b' (that is, both are neither true nor false). But then, how can we express the intuitive difference in meaning between the predicates 'tung' and 'heavy'? The point is that, when we say that two predicates behave in the same way, we do not only mean that they have the same extension and anti-extension, but also that it is necessary that, in case they need to be sharpened, must be sharpened in the same ways. But that means that, if we want to show that two predicates with the same extension and anti-extension behave in different ways, we need to show that they can be sharpened in different ways.

Here Tappenden wants to save one of the motivations for Supervaluationism: the regimentation of the notion of constraint on increases of precision. He uses the supervaluationist machinery in order to define indeterminate sentences, which are those that are true and false depending on the precisification. That means that penumbral sentences are not indeterminate, but have a very special status; they are never appropriately called false, they are never false in any precisification. Nevertheless, both penumbral and indeterminate sentences are neither true nor false (I called them *gappy*).

That is why it can be said that Tappenden follows a position that can be located between the truth-functional and the penumbral intuitions; he concedes to the latter the special status of the penumbral sentences:

> [Penumbral sentences] are never appropriately called false. But contra the penumbral intuition, they are not always correctly called true. (Tappenden 1993, p. 569)

Thus, the general idea that Tappenden has in mind is the following one. Sentences can have two truth values; they can be true (that is, true in the pre-assignment) or false (false in the pre-assignment). On the other hand, they can lack truth value, so that they are neither correctly called true nor correctly called false. Now, among sentences that lack truth value we can distinguish indeterminate sentences (if there is an admissible complete precisification where they are true and an admissible complete precisification where they are false) and penumbral sentences (they are a subset of the pre-analytic sentences and are never false in any complete admissible precisification). Although Tappenden's terminology is rather confusing, I think

this is the best way to make sense of his account. Notice that the use of the supervaluationist machinery is necessary if we want to incorporate somehow the penumbral intuition.

5.4 Some Unsuccessful Objections

5.4 The Penumbral Intuition

Tappenden has to answer an immediate criticism: his approach does not fully respect the penumbral intuition. In his account, as we have seen, a sentence like 'the blob is not both red and orange', where the blob in question is a borderline case of being red and being orange, has to lack truth value, even if we are strongly inclined to regard it as true. The same happens with our first example; a sentence like 'if Joe is taller than John, then, if John is tall, Joe is' lacks truth value when Joe and John are borderline cases of being tall. Thus, we have to answer the question about the reason why we have such a strong intuition.

Tappenden offers the following answer. The use of a language in a given population is a very complex phenomenon. Hence, it is easy that it degenerates in a confusion of tongues. That explains the necessity of maintaining the stability of the conventions of a language by correcting the linguistic mistakes of other people. And that can be made, precisely, using pre-analytic sentences. The idea is that, if you hear somebody that, knowing Joe is taller than John, utters 'John is tall and Joe is not', then you, in order to show her that she has not correctly grasped the use of the word 'tall' and correct her mistake, will utter 'look, if Joe is taller that John, then if John is tall Joe is'. According to Tappenden, the general pattern of this activity is the following one:

> [W]e utter a declarative sentence S in order to induce the withdrawal of a mistaken utterance of $\neg S$, or the withdrawal of other utterances that can only be true if $\neg S$ is true, or to ward off anticipated mistaken utterances of $\neg S$, by indicating that $\neg S$ is never correctly assertable. (Tappenden 1993, p. 570)

Now, since (i) a condition of correctness for a literal assertion of a sentence ϕ is that ϕ must be true and (ii) ϕ is false when $\neg\phi$ is true and, consequently, $\neg\phi$ is not true when ϕ is not false, we can conclude that it is sufficient to show that ϕ is not false in order to show the incorrectness of the assertion of $\neg\phi$. When a sentence ϕ is used in this way to correct a mistaken utterance of $\neg\phi$, we say that ϕ has been articulated, not asserted. The main difference

is that, while assertion implies truth, articulation only implies non falsity. Then, according to Tappenden, penumbral sentences are typically articulated and, therefore, do not need to be true, but only non false —which they are, as they are not false in any precisification. If we sometimes mistakenly judge that they can assert something and, hence, that they need to be true, is because we are confused about assertion and articulation due to the fact that the behavior by which its goals are attained (that is, to say something about the world, and to correct a linguistic mistake) is the same; but it is the same by a happy coincidence. So if the new speech act of articulation is accepted, Tappenden can explain the existence of the penumbral intuition. We will come back to articulation later.

Let's see now two other possible objections to Tappenden's view, one of Rosanna Keefe and another one of Delia Graff Fara, which I think are not successful.

5.4 Articulation and Implicatures

Rosanna Keefe wonders why we could not pragmatically justify a false sentence *in the same way* as Tappenden does. And she continues:

> If I am interested only in preventing assertion or acceptance of false q, and the best way to communicate this is via p because it has the implicature that $\neg q$, then p could be a suitable thing to assert whatever its truth-value. (Keefe 2000, p. 184)

Remember, though, that the kind of process that Tappenden is defending is that, provided that if ϕ is not false, $\neg\phi$ is not true, it is sufficient that ϕ be non false to accept that $\neg\phi$ is not correctly assertable (for a literal assertion of ϕ needs ϕ to be true). That is why the sentence used to correct a linguistic mistake in articulation must be not false in order to imply that its negation is not true (and, then, make its assertion incorrect). There is no implicature here; Keefe is describing something that is not articulation.[11]

Moreover, when Grice (1975) characterizes the notion of conversational implicature, one of its main features, apart from (i) the presupposition of the observance by the speaker of the conversational maxims and (ii) the fact that only the implicature makes sense of a supposed blatant failure of a maxim, is that the speaker thinks and the hearer is supposed to think that the speaker

[11] Maybe I could articulate something that implies ϕ in order to withdraw something equivalent to $\neg\phi$, but it is not still the same thing that Keefe is talking about. Moreover, Tappenden could accept that because he is not claiming that articulation is the only possible way to maintain the stability of languages.

thinks, that the hearer can be aware of the requirement of (ii). Now, in the case Keefe is putting forward, since the maxim to be failed to fulfill is the quality maxim of trying to make the contribution to the conversation one that is true, the speaker utters a false sentence and, then, the hearer must recognize it to be false and must be capable of being aware of a kind of requirement like the one described in (ii) and, only then, an implicature may appear. But in Tappenden's account, we do not only fail to recognize that the articulated pre-analytic sentences should not be called true (and even less we work out condition (ii)); we mistakenly judge them to be true, without realizing that they only need to be not correctly called false in order to succeed.[12] We can see, hence, that it is false that the articulation of a false sentence could be equally pragmatically justified in the same way as Tappenden's.

5.4 The *Sorites*

Delia Graff Fara proposes, in Graff Fara (2000, p. 50), some questions that must be answered in front of the *Sorites* paradox.

In Tappenden's account, the Induction *Sorites* argument is perfectly valid, but the inductive premise lacks truth value (Tappenden (1993, p. 574)), which makes the argument unsound (notice that, within the supervaluationist account, this premise is plain false, for every complete admissible precisification has a sharp limit and, hence, the inference is also incorrect).

Now, according to Graff Fara, there is, first, a semantic question to be answered: if the inductive premise of the paradox, a kind of sentence like $\forall xy((Fx \wedge Rxy) \rightarrow Fy)$[13], is not true, what happens with its classical negation $\exists xy(Fx \wedge Rxy \wedge \neg Fy)$? If it is true, since it seems to deny that vague predicates have borderline cases, we have a problem, for vague predicates have borderline cases. On the other hand, if it is not true, we have a non true sentence with a non true negation, which seems to demand some explanations. Tappenden uses SK in order to defend that the induction premise and its negation are both not true because both lack truth value.

There is also the psychological question: why are we so inclined to accept

[12] Even more; one of the main features of conversational implicatures is that they can be cancelled. Nothing seems to be cancellable in Tappenden's articulation, though. When I articulate a sentence ϕ I utter ϕ in order to induce your withdrawal of $\neg\phi$ and I use the fact that ϕ's non falsity *implies* $\neg\phi$'s non truth. And I cannot cancel an implication.

[13] Recall that F is the vague predicate and R is the tolerance relation.

the conditional premise $\forall xy((Fx \wedge Rxy) \rightarrow Fy)$ if it is not true? Articulation allows Tappenden to solve this question; as we said, we are confused about articulation (that needs only non falsity) and assertion (that needs truth).[14]

Finally, there is what Fara calls the *epistemic question*: if the conditional premise of the paradox $\forall xy((Fx \wedge Rxy) \rightarrow Fy)$ cannot be true, then we should be capable of saying which instances are not true; since it seems that we cannot, an explanation is required. Graff Fara (2000, p. 79) claims, moreover, that Tappenden cannot offer any response to that question.

Tappenden, though, can give an answer to the epistemic question for, according to him, the instances that are not true are those which lack truth value; that is, the ones where either 'Fx' or 'Fy' lack truth value and that, according to the strong Kleene scheme, the result is neither true nor false; that is, the ones where x or y are borderline cases of F (simplifying a bit and supposing that R is not vague).[15] But we can know that. So we can know which instances are not true: the ones that, due to the lack of truth value of its constituents, lack truth value. Thus, if we can know which things are borderline cases of a given predicate (our own response in front of them tells us that) and we can know the semantic rules that govern logical connectives (we know that), we can know which instances of the conditional premise are not true.[16] Hence, we can see that Tappenden can answer Graff Fara's epistemic question.

5.5 Some Objections

Tappenden, on the other hand, criticizes the supervaluationist approach and, indirectly, the necessity of embracing the penumbral intuition. One of the problems for the supervaluationist approach is that, in claiming that the conditional premise of the *Sorites* paradox is false, it is committed to the truth of the claim that there is an n such that n has a certain vague property

[14] Actually, as we will see, things are not so simple here.

[15] To be more precise, according to strong Kleene, the conditional premise will lack truth value (i) when x is a clear case and y a borderline case of F, (ii) when x is a borderline case and y a clear counter-case or (iii) when both are borderline cases (always supposing that Rxy is not vague and true).

[16] I am ignoring higher-order vagueness throughout the chapter. I think that, in general, it is not essential to the points that I am discussing. Nevertheless, it might be important here; if there were higher-order vagueness we could not be capable of deciding whether some objects are borderline cases. But even in this situation we may be able to point out some if not most of the non true instances of the inductive premise, for there will be clear borderline cases.

and its successor does not. But there is not such an n. And a similar thing happens with disjunctions; a disjunction can be true while we are incapable of saying which disjunct is true.

This is one of the main criticisms that Supervaluationism has to face. As Keefe (2000) states:

> One striking departure from classical semantics is the way that there are, in Fine's phrase, 'truth-value shifts', where a disjunction is true though there is no answer to which disjunct is true because the true disjunct shifts from one to another on different [precisifications], or similarly where the true instance of an existentially quantified statement shifts. (Keefe 2000, p. 181)

Supervaluationists have a standard response that can face, at least up to a point, this problem. First, since these truth-value shifts do not appear when we remain within clear cases, and since we already knew that we had to accept something counter-intuitive (the *Sorites* paradox shows that), we can accept this disadvantage because of its role in the whole supervaluationist theory.[17]

[17]It is interesting to consider an argument of Dummett (1975) that is aimed to show, independently of the supervaluationist machinery, that the law of excluded middle is true even if its components are neither true nor false. Suppose we have a vague predicate P and an object a which is a borderline case of P. According to Dummett, it seems plausible to accept that we can find a predicate Q such that it is incompatible with P and such that the sentence *'a* is either P or Q' is true. Now, since P and Q are incompatible, Q implies not P and, hence, whenever *'a* is either P or Q' is true, *'a* is either P or not P' is also true. Using this idea Dummett claims that it is reasonable to say that, for any sentence A involving vague expressions, and independently of its truth value, the sentence of the form *'A* or not A' may be seen as true. That is why he says: 'when vague statements are involved, then, we may legitimately assert a disjunctive statement without allowing that there is any determinate answer to the question which of the disjuncts is true' (Dummett (1975, p. 107)). Moreover, this argument could prompt the suspicion that sentences involving vague predicates behave classically and, hence, it could be seen as a reason in favor of any approach that tries to respect classical logic (*v. gr.* Supervaluationism or some epistemicist approaches that keep classical logic, like Horwich's approach we are going to see in Chapter 6) . It remains to be seen, though, if the argument succeeds; after all, if we want the argument not to be question-begging we need to find, given a vague predicate P and a borderline case a of P, another predicate Q incompatible with P such that a *is not* a borderline case of Q, which does not seem that plausible. That is, if we want to show that (a) 'Paul is tall or Paul is not tall' is true even when Paul is a borderline case of

Second, Supervaluationism can distinguish between the truth of the negation of the inductive premise of the *Sorites* paradox (that is, that there is an n that has a certain property and its successor does not) and its having a true instance. That can explain, claims the supervaluationist, our mistaken intuitions.

The idea is that when we accept the negation of the inductive premise we are not actually accepting the existence of a sharp boundary. The problem is that we get confused between the claim that (i) it is true that, for some n, n has a property and n+1 does not, and the claim that (ii) for some n, it is true that n has a property and it is false that n+1 does. These two claims do not need to have the same truth value; the second can be false while the first is true. The confusion is a confusion of the scope of the truth predicate; when it appears outside the existential quantifier, the resulting sentence can be true without an instance making it true. Now we can see that in the objection above we are confusing (i) and (ii).

The problems for the supervaluationist, though, do not end here. We have seen that the precisifications over which she quantifies in order to evaluate sentences must be complete; that means that it is necessary to draw sharp boundaries between the extension and the anti-extension of vague predicates. Tappenden claims that there is a certain kind of predicates, the essentially vague ones, whose understanding implies the impossibility of drawing such sharp boundaries. He defines essentially vague predicates in the following way:

> Call a predicate P essentially vague if there is a sequence a_1, a_2, ..., a_n, and a relation Q, such that a_1 is a clear case of P, a_n is a clear counter-case, for each i, n+1, Qa_ia_{i+1} is true, and each instance of 'If Qa_ia_{i+1} then (Pa_i if and only if Pa_{i+1})' is a local consistency rule.

where a local consistency rule is a pre-analytic sentence, that is, according to Tappenden,

> a sentence which anyone who understands [it] knows not to draw more precise boundaries [to P (supposing that Q is not vague)] in such a way that [the sentence] would be false in any circumstances. (Tappenden 1993, p. 557)

being bald, it is not enough to appeal to the sentence (b) 'Paul is tall or Paul is short'. Although it is true that being short implies being non tall and, that, hence, if (b) is true so is (a), the problem is that, if Paul is a borderline case of being tall he will also be a borderline case of being short and, hence, someone who defends that (a) is not true will very likely think that (b) is not true either.

That means that there are predicates (essentially vague predicates like 'roughly heavy', 'roughly within walking distance of Barcelona' or 'roughly a handful of sand') which do not accept complete precisifications in virtue of its meaning.[18]

This criticism is very close to one presented by Michael Dummett. He claims that, although vagueness somehow invests language with intrinsic incoherence, it is also an essential feature of language. Then, the problem with Supervaluationism is that it regards vagueness as if it were eliminable and, thus, it does not take vagueness seriously enough; it could seem that, according to Supervaluationism, the fact that our language is vague is just due to our laziness to make it precise.

The supervaluationist can respond that, after all, her theory is a semantic account that quantifies over precisifications and that it is this quantification what tries to capture the meaning of vague predicates, not the individual precisifications. That means that it does not matter if it is impossible to use in practice one of the precisifications (for example, due to the fact that our language is essentially vague).

Nevertheless, the criticism of essentially vague predicates is more worrisome for Supervaluationism. After all, we need to precisify vague predicates in order to evaluate vague sentences; we may do that without committing ourselves to the entities over which we quantify, but we cannot do that if we are not able to give the rules that constrain such precisifications. And that is what happens with predicates whose meaning entails a consistency rule; the very meaning of the predicate prevents us from drawing sharp boundaries. That means that the set of all complete admissible precisifications is empty and that, consequently, we cannot evaluate any vague sentence (as a matter of fact, all of them would be true and false).

In more detail, the worry is the following one. When we are in front of an essentially vague predicate, we have sentences (local consistency rules) of the form $\forall xy(Rxy \rightarrow (Px \leftrightarrow Py))$ that, in virtue of the meaning of the predicate

[18]Other authors like, for example, Eklund (2001) propose similar predicates. According to Eklund, predicates like 'smallish' or 'roughly red' present a serious problem to Supervaluationism because they challenge the rationale for singling out a particular set of precisifications as acceptable: "according to the supervaluationist analysis, vagueness is a matter of what we have failed to lay down. But it appears that the conventions associated with 'roughly' and '-ish' say exactly that constructions with these expressions are supposed to be vague"(Eklund 2001, p. 366). Eklund seems to consider that problem as unsolvable and, hence, the only possibility for Supervaluationism is to treat such predicates as special cases to which precisifications cannot be applied. That means that then Supervaluationism cannot offer a unified account of the phenomenon of vagueness.

P cannot be false; that is, they are pre-analytic according to Tappenden. In contrast, suppose that P is a non essentially vague predicate, say, 'tall', and that Rxy if and only if y is one millimeter taller than x. Then, one of the directions of the biconditional, namely (i) $\forall xy(Rxy \rightarrow (Px \rightarrow Py))$, is clearly pre-analytic in Tappenden's sense. But the other direction, (ii) $\forall xy(Rxy \rightarrow (Py \rightarrow Px))$ is not (that is why it can be false within the supervaluationist frame). The latter is pre-analytic, though, if P is 'roughly tall'; that is, (ii) is never false, and that happens in virtue of the very meaning of P.

Notice that (ii) is the conditional premise of the *Sorites* paradox. And we are strongly inclined to consider it as true —if we were not, we would not have a paradox. Why? Well, when P is a non essentially vague predicate like 'tall' we would say that the very meaning of the predicate P seems to suggest that it is true. But, that is precisely what happens with (ii) when P is an essentially vague predicate like 'roughly tall'. The predicates 'tall' and 'roughly tall' are certainly different but they seem to provide (ii) with the same semantic status; in both cases (ii) seems true in virtue of its vague expressions. Hence, it is not clear at all in what could consist the difference between essentially and non essentially vague predicates and in what sense their supposed differences in meaning could prevent sharp boundaries from being drawn in the former case and not in the latter. Let us help illuminate the situation with an example.

Take the predicate 'roughly tall', why cannot we precisify this predicate saying that it increases the extension of the predicate 'tall' in, say, three centimeters? The idea is, then, that there will be a set of admissible precisifications of the first predicate that will extent the extension of the second predicate. That means, of course, that somebody of two meters will be roughly tall; that could sound weird, but it does not seem very difficult to give a plausible pragmatic story capable of explaining such a weirdness: as Matti Eklund proposes, when we say that Goliath is roughly tall, we are flouting the Gricean maxim *Be specific*, but we are not saying anything untrue (Eklund 2001, p. 366).

In light of these considerations, we conclude that there is no difference between essentially and non essentially vague predicates; both allow precisifications and, more importantly, both are governed by the same pre-analytic sentences. Hence, Tappenden has two options; first, he can accept that there are no essentially vague predicates, in which case his objection to Supervaluationism based on this kind of predicates cannot get off the ground or, second, he can claim that all vague predicates are essentially vague predicates. Then, his criticism collapses into Dummett's one that precisifying vague predicates does not respect their meaning. And we have seen that Supervaluationism

has a response to Dummett's worry.

Now the supervaluationist must say that (ii) is false even when the predicate is essentially vague; I do not see why that is harder to accept than the cases where the predicate is non essentially vague; denying (ii) is equally weird (if weird at all) independently of the degree of vagueness of the predicate.

Besides, the situation does not change if we consider precisifications as primitive; after all, if we can stipulate, within a great amount of arbitrariness, sharp boundaries for the predicate 'tall' it is not clear why we cannot do the same with the predicate 'roughly tall' (recall that the supervaluationist does not need to commit herself to the use in practice of any of the precisifications). Thus, finally, the supervaluationist can claim that there are not essentially vague predicates and that Tappenden's local consistency rules are not only no pre-analytic, but false.

Consequently, if we do not accept the existence of essentially vague predicates, Tappenden's argument against Supervaluationism fails. Additionally, we will see now that if we accept the existence of essentially vague predicates, not only the supervaluationist has to face a serious problem, but also has Tappenden.

Recall that the supervaluationist machinery was essential to Tappenden's approach. So if it is true that the set of complete admissible precisifications is empty, Tappenden also has a problem for, how can he distinguish between the predicates 'roughly tung' and 'roughly heavy?' They behave in very similar ways but, without the supervaluationist machinery, it seems difficult to express the ways in which they differ. Now Tappenden cannot use the set of complete admissible precisifications in order to differentiate indeterminate sentences from penumbral sentences and then, how can he distinguish them? They are sentences with exactly the same status: neither true nor false. Thus Tappenden, in accepting the existence of essentially vague predicates, is refuting his own point of view, or at least he is weakening it.

Hence, even if we suppose that there are predicates that cannot be sharpened, Tappenden's view is, at least, as problematic as Supervaluationism.

Another noticeable feature of Tappenden's account is related to the reason why the deception of the paradoxes is so compelling, which, as we saw in chapter 3, should be part of any solution to the paradoxes. As we have seen, articulation is what allows Tappenden to solve this question; we are confused about articulation (that needs only non falsity) and assertion (that needs truth). But we can see now that this proposal works only in the case of essentially vague predicates, where the main premise of the Induction *Sorites* is a pre-analytic sentence, that is, it is never false in any precisification. In contrast, this is not the case when the predicate is not essentially vague,

for then, the inductive premise of the Induction *Sorites* is not pre-analytic, for it is false in some precisifcations, which means that it is indeterminate. The problem is that only the pre-analytic sentences can be articulated and, hence, the main premise of the Induction *Sorites* cannot be articulated, which implies that we cannot explain why it seems true to us. Tappenden's response is that, since in most contexts we do not want to draw boundaries sharply enough for the sentence to be false, it seems true to us. This suggests that precisifications might not be complete; this, tough, means that, either we must give up the supervaluationist machinery, which, as I said is a serious problem for Tappenden, or we must find another way of evaluating sentences in non complete precisifications. I will come back to this last point in the next section.

There is still another problem with Tappenden's use of articulation. He says that what explains the fact that we tend to see penumbral sentences as true is that we confuse assertion and articulation. Thus, if we believe that the sentence 'it is not the case that the blob is red and the blob is not red' uttered in front of a borderline case of being red is true is due to the fact that we typically use it in situations were we think it has to be true (we think that we are asserting it) but, actually, it has to be only non false (we are articulating it). It seems true that, if we imagine a situation where somebody utters 'this blob is red and not red', then, in order to correct her we could say 'it is not the case that the blob is red and not red' and, moreover, that this latter sentence only need to be non false in order to succeed. But can we really rely on this kind of situations in order to justify the existence of the very strong intuition that penumbral sentences are true? The confusion between assertion and articulation that helps Tappenden support the truth-functional intuition seems to rely on the fact that we have to be in situations like these very often, sufficiently often to explain our mistake. But that does not seem to be the case; most of us are confronted to sentences like 'if Joe is taller than John, then, if John is tall, Joe is' for the first time when we begin to read papers about vagueness and, in spite of that, we are strongly inclined to believe them true. I do not see any room for confusion here.

That suggests that offering an account of the penumbral intuition is still a problem for Tappenden's approach to vagueness.

5.6 The Liar

As we said at the beggining, Tappenden accepts Kripke's framework in order to cope with the Liar. We also saw two of the main problems this approach has to face. Let us put the first one in another way.

Recall that the Liar reasoning allows us to conclude that the Liar is true if, and only if, it is not true. Now imagine yourself explaining this paradox to someone who does not seem to understand it. As Tappenden observes, in a situation like that,

> [y]ou might well say: "Here is what is funny about [the Liar]. If [it] is true, then it is not true, and if [it] is not true then it is true". (Tappenden 1993, p. 552)

Although this might seem contradictory, it is the appropriate thing to say in this situation; we seem to be asserting it appropriately. But as we saw at the beginning, the sentence $Tr\ulcorner\lambda\urcorner \leftrightarrow \neg Tr\ulcorner\lambda\urcorner$ (the formal counterpart of the previous claim) has value ½ in Kripke's construction, what implies that it should not be assertable. This is just a consequence of the weakness of the conditional in *SK* we already discussed above.

Now, Tappenden's strategy consists in defending that "some of the considerations that pertained to vague predicates may be carried over" (Tappenden 1993, p. 575) to the truth predicate, so that articulation can be used in order to explain away why certain sentences that have no truth value according to Kripke's approach seem true to us; that is, Tappenden wants to explain away the penumbral intuition applied to the truth predicate. Some of the sentences Tappenden thinks fall under the penumbral intuition are, apart from the already noticed $Tr\ulcorner\lambda\urcorner \leftrightarrow \neg Tr\ulcorner\lambda\urcorner$, instances of LEM applied to ungrounded sentences and, of course, the instances of the T-schema of ungrounded sentences. All these sentences are pre-analytic; that is, they are never false, in a sense to be specified shortly. Besides, since they receive value ½ in *SK* they are penumbral sentences.

Let us see, first, which are the features of the truth predicate that make Tappenden claim that we can carry over to truth the results we got about vagueness. The main such feature Tappenden puts forward is the arbitrariness that can be found when we try to determine which is the extension of the truth predicate. Specifically, Tappenden mentions the Truth-teller, the sentence τ introduced just at the beginning of this chapter that is identical to its own truth ascription. As we saw, and so Tappenden claims, we can consistently assign τ to the extension of Tr and we can also consistently assign $\neg\tau$ to this extension. This suggests, according to him, that there will exist constraints on how these different stipulations of the extension of the truth predicate might be. Among these constraints there will be sentences like 'for any sentence x, $Tr\ulcorner x\urcorner \vee \neg Tr\ulcorner x\urcorner$' or, indeed, any sentence of the form $Tr\ulcorner\phi\urcorner \leftrightarrow \phi$. These sentences are evaluated as having semantic value ½ in Kripke's framework and, hence, they are not assertable. Why, then, in situations like the one described above, we seem to assert them? Why do they

seem true to us? How can we explain away the penumbral intuition applied to truth? Here is where Tappenden uses the notion of articulation:

> As with vague predicates, we may explain away the penumbral intuition by noting the way the patterns of use of the relevant sentences lead us to take them unreflectively to be true. The Tarski biconditional with liar instances [...] might well be uttered in the course of an attempt to demonstrate why it is unacceptable to assert [the Liar] and unacceptable to assert the negation of [the Liar]. The imagined utterance of Tarski biconditional is successful if the hearer recognizes that certain other sentences cannot be correctly asserted; so the account of articulation [...] extends naturally to this expanded setting. (Tappenden 1993, p. 576)

Recall now how articulation worked: essentially, we uttered a sentence ϕ in order to show that it was unacceptable to assert $\neg\phi$. But since in order to show that $\neg\phi$ cannot be asserted it is enough to show that it is not true, and since ϕ being not false implies that $\neg\phi$ is not true, we concluded that showing that ϕ is not false is enough to show that $\neg\phi$ cannot be asserted. Now, in the case of vagueness, pre-analytic sentences were typically articulated, according to Tappenden, because they were never false in any precisification. Can we adapt this idea to truth? I do not think so.

As I said at the presentation of Kripke's approach, $\boldsymbol{K}$ is just the minimal fixed point; there are many other fixed points that extend $\boldsymbol{K}$ and that, as fixed points, validate the Principle of Intersubstitutivity (so that they can be claimed to be a possible extension for a truth predicate). Kripke (1975) also showed that every fixed point can be extended to a maximal fixed point, where a maximal fixed point, in Kripke's words, is "a fixed point that has no proper extension that is also a fixed point. Maximal fixed points assign "as many truth values as possible"; one could not assign more consistently with the intuitive concept of truth" (Kripke 1975, p. 708).

Now, if Tappenden's use of articulation applied to the Liar has to make sense, the instances of the T-schema will be pre-analytic sentences; that is, sentences that cannot be evaluated as false in any way of making more precise the truth predicate. What that means is that the most natural candidates to be the admissible ways to make more precise the truth predicate are, precisely, the fixed points, because they never make false any instance of the T-schema. Recall, though, that, apart from being admissible, the precisifications of the vague predicates had to be complete; all cases had to be decided. In the case of truth, this is not possible, if the instances of the T-schema are regarded as pre-analytic. To see why, suppose X is a complete way of making the truth predicate precise and N is a model for the base language

$\mathcal{L}$ as introduced in section 5.1. Now, if X is complete, either $|\lambda|_{\langle \mathcal{N},X\rangle} = 1$ or $|\lambda|_{\langle \mathcal{N},X\rangle} = 0$. Suppose $|\lambda|_{\langle \mathcal{N},X\rangle} = 1$ (the other case is analogous). Then, by definition of λ, $|\neg Tr^{\ulcorner}\lambda^{\urcorner}|_{\langle \mathcal{N},X\rangle} = 1$ and, hence, $|Tr^{\ulcorner}\lambda^{\urcorner}|_{\langle \mathcal{N},X\rangle} = 0$. Consequently, $|Tr^{\ulcorner}\lambda^{\urcorner} \leftrightarrow \lambda|_{\langle \mathcal{N},X\rangle} = 0$. So, if the truth predicate is made completely precise some instances of the T-schema will be false and, hence, the T-schema cannot be pre-analytic. That means that the ways of making precise the truth predicate cannot be complete, which, in turn, means that classical logic cannot be used in them, in sharp contrast with vagueness. It is to be expected, then, that in order to evaluate the semantic value of the sentences in the fixed points, we will have to use *SK*.

Hence, the picture we are unraveling is the following one. The semantic value of the sentences involving truth is determined by $\boldsymbol{K}$, the minimal fixed point. Then some sentences that have no truth value are such that we are strongly inclined to believe them true. This inclination, though, is an illusion prompted by the fact that, although they do not have truth value, they cannot be false, which is what is needed to correct linguistic mistakes. Now, since given that they are never false and, hence, they are used to correct linguistic mistakes and given that correcting linguistic mistakes with such sentences (articulating them) is so similar to asserting them, we confusingly conclude that they are asserted, which in turn implies that they are true. In order to make sense of the idea of a sentence with vague predicates that cannot be false, Tappenden used the Supervaluational machinery. Then, the suggestion is to make a similar move for the truth predicate; pre-analytic sentences involving truth will be the sentences that are not false in any way of making precise the truth predicate that does not conflict with its meaning, that is, the sentences that are not false in any fixed point. But the fact that the ways of making the truth predicate more precise are not complete is fatal for, although it is true that sentences like (i) 'for any sentence x, $Tr^{\ulcorner}x^{\urcorner} \vee \neg Tr^{\ulcorner}x^{\urcorner}$' are never false in any fixed point, the same happens, for example, to (ii) 'for some sentence x, $Tr^{\ulcorner}x^{\urcorner} \wedge \neg Tr^{\ulcorner}x^{\urcorner}$'. This means that if we characterize pre-analytic sentences as the sentences that are never false in any way of making 'true' precise, (ii) will be pre-analytic. But pre-analytic sentences were supposed to capture some important features of the meaning of the predicates involving them, which in the case of (ii) is clearly not the case. This means that Tappenden should provide us with a new characterization of pre-analytic sentences that is not clear at all how could proceed.

The truth predicate, thus, becomes something on the lines of an essentially vague predicate; there are no complete and admissible ways of making the truth predicate precise. That makes difficult to imagine how can we apply the strategy that Tappenden used for vague predicates to the truth predicate.

5.7 Solving the paradoxes

It remains to be seen whether Tappenden is defending something on the lines of a common solution to the Liar and the *Sorites*. Recall that in chapter 3 we defended that a common solution to both paradoxes must offer at least a common reason about why there is a deception in them. We also distinguished between a strong and a weak common solution to the Liar and the *Sorites*; the former offers the same prevention to both paradoxes while the latter does not.

In this chapter we have seen that what Tappenden seems to propose is the following:

(a) First, he offers preventions to the Liar (Kripke's proposal) and the *Sorites* (*SK*) which have in common the use of *SK*. Still, even if both preventions use *SK*, they are not the same; all the sophisticated machinery about fixed points used in the case of truth does not play any role in the case of vagueness.

(b) Second, Tappenden offers a common explanation of the reason why the deception in the paradoxes (the non-truth of the inductive premise in the case of the *Sorites* and the invalidity of LEM in the case of the Liar) is compelling, which in Tapenden's framework is equivalent to explaining away the penumbral intuition. As we have seen, Tappenden uses ideas of Supervaluationism and a new speech act, articulation, in order to do that. We have also seen that, unfortunately, it is not clear at all whether this strategy can be soundly applied to the truth predicate.

Does any of (a) or (b) above imply that we are in front of a common solution to the Liar and the *Sorites*? Having in mind the previous considerations the answer is no; no common reason about why there is a deception in the paradox is offered.

Nevertheless, Tappenden's proposal, if successful, would be a tool that could be used by certain common solution to the Liar and the *Sorites*. Imagine we have a solution to both paradoxes that proposes that the Liar and borderline vague ascriptions are undetermined and that points to a common source of this indeterminacy, so that such a solution can be considered a common solution. Suppose, further, that some logic similar enough to *SK* was used in the preventions of both paradoxes so that, due to its weakness, the penumbral intuition remained unanswered. In cases like these, Tappenden's articulation might be used in order to explain away this intuition, what amounts to explaining away why the culprits of the Liar and the *Sorites*

paradoxes are so compelling; that is, why the T-schema (and LEM) and the inductive premise seem true to us.

CHAPTER

SIX

PAUL HORWICH: SEMANTIC EPISTEMICISM

Paul Horwich defends an epistemic account for vagueness and its paradoxes. He and other authors have tried to look into the possibility of applying the strategy Horwich devises for vagueness to truth. In this chapter I will present Horwich's approach to vagueness, which preserves classical logic, and I will discuss whether and how it can be applied to the case of the Liar.

I will need to introduce Horwich's deflationary theory of truth, called *Minimalism*, and present the solution Horwich offers for the Liar. It will turn out that such a solution proceeds by restricting the T-schema and, as a consequence of that, it will need a constructive specification of which instances of the T-schema are to be excluded from the minimalist theory of truth. Horwich has presented, in a very informal way, one such construction that would specify the minimalist theory. In the last part of the chapter I will try to make it precise. It will turn out that the construction is the minimal fixed point of Kripke's construction with the supervaluationist scheme. Finally, some properties of Horwich's construction and some amendments to it will be discussed.

6.1 Vagueness

Horwich defends an epistemic account of vagueness according to which vague predicates have sharp boundaries which we are not capable of knowing. Furthermore, he wants to preserve the Law of Excluded Middle (LEM), which is seen as a basic law of thought, and, consequently, he claims, the Principle of Bivalence, or BIV (see, for example, Horwich 1997, 2005a). He needs, hence, a theory of vagueness capable of accommodating all these features and of explaining the phenomenology that underlies the phenomenon of vagueness;

that is, according to him, our tendency to be unwilling to apply both the predicate and its negation to certain objects although being aware of the fact that no further investigation could be of any usefulness.

Horwich admits that the following claim is counterintuitive:

(*) vague predicates have sharp boundaries, that is, they divide the world into two sharp groups of things; the ones that have the property expressed by the predicate and the ones that haven't.

He claims, though, that in front of the *Sorites* paradox only two reasonable responses are possible: abandon classical logic or accepting (*). Since the former strategy is seen by Horwich as desperate, he proposes to follow the latter one: we must accept (*). But the intuition that vagueness is at odds with sharp boundaries is very strong and, if we try to solve the Sorites paradox by posing sharp boundaries on vague predicates, we still need to explain away why we have such a strong intuition. One of the main roots of our reluctance to the acceptance of sharp boundaries for vague predicates is the fact that it seems impossible to find them; it seems impossible to find out the line that divides the objects that have a given property expressed by a vague predicate and the objects that do not have that property. But the fact that we are not able to find out the sharp boundaries of vague predicates can be best explained by the fact that there are no such sharp boundaries. Horwich's response to this line of thought consists of an explanation of why we cannot know where the sharp boundaries of our vague predicates are and, consequently, why we can't know the extensions of such predicates. Let's see how this account is articulated.

Horwich proposes to look at the fundamental regularities implicit in our linguistic practice that underly our use of vague predicates. His proposal is to understand this fundamental regularity as

> approximated by a partial function [...] which specifies the subjective probability of its applying as a function of the underlying parameter n (i.e. 'number of grains' for 'heap', 'number of dollars' for 'rich',...). (Horwich 1997, p. 933)

Such regularities would explain all our uses of vague predicates; they would be complete in the sense that any "decision" (Horwich 1997, p. 934) about the borderline cases of a given vague predicate P would have to be a consequence of the underlying partial function *A(P)*. Such function would have been implicitly acquired by exposure to sentences reflecting clear instances of P, of not-P, and of not so clear cases close enough to the clear ones. The partial function *A(P)* is determined, thus, by our acceptance of certain sentences containing the word for P. That's why Horwich says:

> the explanatorily fundamental acceptance property underlying our use of 'red' is (roughly) the disposition to apply 'red' to an observed surface when and only when it is clearly red. (Horwich 1998a, p. 45)

These partial functions are the fundamental facts about our use of vague predicates; they are fundamental in the sense that they must be in the basis of any explanation of any fact concerning our use of vague predicates (Horwich 2005a, p. 94, 1997, p. 934). The fact that these fundamental facts are functions that remain silent with respect to the application of the predicates to certain objects explains why we also must remain silent in front of such applications and why we are confident that acquiring new information will not solve the matter, which is, according to Horwich, the basic phenomenology underlying vagueness.

These considerations also explain why we cannot know whether borderline cases are in the extensions of vague predicates; the only way we can be justified in applying a given vague predicate to an object is via the fundamental facts underlying the vague predicate and that, as we have seen, is not possible. Hence, believing an ascription of a vague predicate to a borderline case will never be able to constitute knowledge.

Horwich then defines a notion of determinateness based on his account of meaning. He uses, first, the claim that meanings are concepts and, second, that meaning properties are constituted[1] by use properties, which, in turn, stem from some given fundamental acceptance properties (Horwich 1998a, p. 44). The idea, then, is that some ascription of a predicate P to an object o is indeterminate when it is conceptually impossible to know whether o is P; that is, when the unknowability of the ascription has its roots in the facts (that is, the fundamental acceptance properties) that make our words mean what they mean.

I will not focus here on the virtues or defects of Horwich's account of vagueness, but rather, I will investigate whether this approach can be applied to truth and the Liar, as Horwich himself seems to defend.

[1] According to Horwich (1998a, p. 25) a given property A is constituted by another property B when their coextensiveness is the basic explanation of facts involving A; thus, for example, if the property of being water is constituted by the property of being composed of H_2O, it is because (i) they apply to the same things and (ii) that this fact (namely, (i)) explains all facts about the property of being water.

6.2 Minimalism and Semantic Epistemicism

6.2 Deflationism

Horwich's theory of truth, called *Minimalism*, follows the Wittgensteinian rule against overdrawing linguistic analogies; although for some predicates ('table', 'dog',...) it makes sense to inquire into the shared characteristics of the things to which they apply, for some others, like the truth predicate, it does not. If it makes sense to seek some kind of underlying nature in the case of the former kind of predicates, it is because they are used to categorize reality; we cannot presuppose, though, that this is the function of the truth predicate. Actually, Horwich is a deflationist about truth, which means that, according to him, truth is not a genuine property; the truth predicate is not used to describe anything, the true truth-bearers do not share any common property. As a matter of fact, the truth predicate is just a device of disquotation. Truth, hence, is a semantical, or logical, notion that is in no need of metaphysical or epistemological analysis. A *locus classicus* for the view that the truth predicate is just a device of disquotation that enables us to express, by means of quantification, certain infinite conjunctions and disjunctions is Quine (1986):

> We may affirm the single sentence by just uttering it, unaided by quotation or by the truth predicate; but if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent. (Quine 1986, p. 12)

Why would we need to express such infinite lots of sentences? Suppose I, for some reason, believe all you said yesterday. I can express this belief using the following infinite conjunction, where each ϕ_n is to be replaced for a sentence in the language:

> (If you said yesterday: 'ϕ_1', then ϕ_1) and (If you said yesterday: 'ϕ_2', then ϕ_2) and ...

Now, since we cannot handle infinite conjunctions, it is tempting to directly generalize on ϕ's position above using ordinary pronominal variables, and obtain:

> For every x, if you said yesterday: 'x', then x.

This will not work, though, because on the one hand, the second occurrence of the variable x above is in an opaque context and, thus, cannot be bound by usual quantifiers and, on the other hand, pronominal variables can be substituted only by singular terms and the third occurrence of the variable appears in a sentence position and, thus, it cannot be substituted by a singular term.

We can easily solve the problem for the second occurrence of the variable in the last generalization and get:

> For every x, if you said yesterday: x, then _.

Where the second occurrence of x is used now in a name position (the name of the sentence uttered). If we could do the same with respect to the third occurrence of the variable, we would have reached the desired goal. And that is what the truth predicate allows us to do. The key idea is that a sentence ϕ and the sentence that says that ϕ is true are always interchangeable *salva veritate* and, hence, the truth predicate, via its disquotational character, allows us to transform a given sentence ϕ into another sentence which makes an ascription of a predicate (the truth predicate) using a name for ϕ and, thus, allows us to use ordinary pronominal variables:

> For every x, if you said yesterday: x, then x is true.[2]

Or, plainly

> All you said yesterday is true.

As Gupta (1993a, p. 61) puts it, the truth predicate "enables us to generalize over sentence positions *while using pronominal variables such as 'x'* and, thus, endows us with additional expressive power".

[2]There is another way to circumvent this problem; we can use substitutional quantification. Substitutional quantifiers can bind variables of an arbitrary substitution class and hence, in particular, they can bind sentence variables and predicate variables. According to Horwich, though, the advantage of the truth predicate over the substitutional quantification is that the former is a simpler linguistic device than the latter, which would be a "cumbersome addition to our language" (Horwich 1998b, p. 32). I do not think, though, that a deflationist needs to commit herself to claims about the higher efficiency of the truth predicate over other similar linguistic devices. The truth predicate is the device we have in natural languages to overcome certain expressive limitations but, as fas as I can see, the proponent of the deflationary point of view about truth does not need to defend that it is the only device (not even the best one) that can do that.

6.2 Minimalism

This was a general explanation of why deflationists about truth think we have the truth predicate in our language. But different deflationary approaches to truth develop these intuitions in different ways. Thus, for example, although some deflationists think that the truth-bearers are utterances (see, for example Field 1994) some others, like Horwich himself, think that the truth-bearers are propositions. Besides, the idea of the interchangeability of a sentence and its truth ascription is usually captured by deflationists with the T-schema, which in the case of Horwich, is applied, as I said, to propositions:[3]

(T-schema) $< p >$ is true iff p.

Horwich (1998b, 2001, 2010c) has presented and defended Minimalism. One of its main theses is that the instances of the T-schema are epistemologically, explanatory and conceptually fundamental. Thus, in the first place, they fix the meaning, they implicitly define the truth predicate (Horwich 1998b, p. 145); this is so because the basic and fundamental regularity of use that determines the meaning of 'truth' (which is the concept of truth, for meanings are concepts, according to Horwich) is our disposition to accept all instances of the T-schema, so they are conceptually fundamental. In the second place, the instances of the T-schema are all we need to explain all our uses of 'true', so they are explanatory fundamental.[4] And, finally, the instances of the T-schema are "immediately known" (Horwich 2010c, p. 36), they cannot be deduced from anything more basic, so they are epistemologically fundamental. In other words, according to Horwich, the role of the T-schema with respect to the truth predicate is the same as the partial functions of section 6.1 with respect to vague predicates.

Considering all that, therefore, it is not surprising that Horwich's theory of truth, Minimalism, contains as axioms all instances of the T-schema applied to propositions, and nothing else.[5]

[3]The symbols '<' and '>' surrounding a given expression e produce an expression referring to the propositional constituent expressed by e. Thus, when e is a sentence, '$< e >$' means *the proposition that e*.

[4]This is an exaggeration; strictly speaking, we will need other theories besides the truth theory to explain all facts about truth, because some of these facts will involve other phenomena. As Horwich says, Minimalism "provides a theory of truth that is a theory of nothing else, but which is sufficient, in combination with theories of other phenomena, to explain all the facts about truth" (Horwich 1998b, pp. 24-25).

[5]That characterization is not completely accurate; as Horwich admits, the theory should also have an axiom claiming that only propositions are bearers of truth (see

Now, as we know, the proposition that asserts its own untruth makes the theory consisting of just all instances of the T-schema inconsistent in classical logic. Until recently, Horwich's response to this problem had been very succinct. In his (1998) he claims that the lesson the Liar tells us is that not all the instances of the T-schema are to be included as axioms in the theory (Horwich 1998b, p. 42). Thus, the minimalist theory of truth consists of a restricted collection of instances of the T-schema; only those that do not engender Liar-like paradoxes. Which of the instances of the T-schema should be removed, though, was left undetermined.

A proposal of a full solution to the Liar based on the previous considerations has been made explicit by Armour-Garb (2004), Beall and Armour-Garb (2005), Restall (2005) and, though succinctly, by Horwich himself in his (2010). Beall, Armour-Garb and Restall has called it *Semantic Epistemicism*. Horwich claims:

> [W]e can and should preserve the full generality of the Law of Excluded Middle and the Principle of Bivalence: [The Liar] is either true or false. Of course we cannot come to know which of these truth values it has. For confidence one way or the other is precluded by the meaning of the word 'true' — more specifically, by the fact that its use is governed by the [T-schema] (subject to the above restrictions). Thus, just as it is indeterminate whether a certain vague predicate applies, or does not apply, to a certain borderline case (although certainly it does or doesn't), so (*and for the same reason*) it is indeterminate whether [The Liar] is true or whether it is false. (The emphasis is mine)(Horwich 2010c, fn. 11 in page 91)

We can now present the two tenets of Semantic Epistemicism, the minimalist stance in front of The Liar:

1. The Liar is true or The Liar is false.

2. It is conceptually impossible to know whether The Liar is true and it is conceptually impossible to know whether The Liar is false.

Let's see the rationales for these two points.

First Tenet First, 1 is an instance of the Principle of Bivalence. Horwich (1998b, p. 71) justifies this principle applied to all propositions in the following way. Define, first, falsity in terms of "absence of truth":

> $< p >$ is false iff it is not the case that $< p >$ is true.

Horwich 1998b, fn. 7 in page 23, page 43).

Then, the Law of Excluded Middle gives us the desired result. For given a proposition p we have that p or not-p (LEM) and, hence, in particular when p is of the form *q is true* , we obtain that either q is true or it is not the case that q is true; that is, either q is true or q is false. Notice that we did not need to use the instance of the T-schema for p. The problem is that Horwich seems to be begging the question when he defines falsity as non-truth, because he is supposing that the only way in which a proposition can be not true is by being false; but that is, precisely, what is at stake here. Consider, for example, a view that posits truth value gaps (like, for example, some form of Supervaluationism); such a view will not accept Horwich's definition of falsity, because propositions that are neither true nor false are, in particular, not true, but they are not false either; so falsity cannot be absence of truth.

A more neutral way of defining falsity is in terms of truth of the negation:

$< p >$ is false iff <not $p >$ is true.

The problem is that, if Horwich had defined falsity in terms of truth of the negation rather than in terms of non-truth, then he would have needed to use the instance of the T-schema for p and, since some of the instances of the T-schema are not in the theory (in particular the Liar instance) he would not have been able to derive 1. So Horwich seems to be in front of a dilemma; either he does not use the T-schema but begs the question, or he offers a more dialectically robust argument which turns out to be unsound due to the restriction on the T-schema. [6]We will see, towards the end of this chapter, another way to obtain bivalence for the Minimalist theory.

Second Tenet The reasons for accepting 2 are closely related to our previous discussion of vagueness. As Horwich says in the quote above, the reasons why the truth value of the Liar is indeterminate are the same as in the case of vagueness. Hence, the reasons for accepting 2 are rooted in the fact that the instances of the T-schema are explanatorily fundamental in the sense that they must be in the basis of any explanation of any fact concerning our use of the truth predicate and, moreover, that they fix the meaning of 'truth', that is, the concept of truth.

As in the vagueness case, Horwich can define a notion of determinateness based on his account of meaning. He uses, first, the claim that meanings are concepts and, second, that meaning properties are constituted by use

[6]Even with the T-schema unrestricted, it is contentious that BIV can be vindicated using LEM, specially when the T-schema is formulated in a congenial way to Supervaluationism. See Andjelcović and Williamson (2000) and López de Sa (2009).

properties, which, in turn, stem from some given fundamental acceptance properties (Horwich 1998a, p. 44), which, in the case of the truth predicate, are our dispositions to accept the instances of the T-schema. Some of these dispositions, though, are overridden by the fact that we realize that some of the instances of the T-schema lead to inconsistency. These instances, then, will not be in the minimalist truth theory.

The idea, once we have restricted the truth theory, is that some ascription of the truth predicate to a given proposition p is indeterminate when we do not know whether p is true or not and the unknowability of the ascription has its roots in the facts (that is, the fundamental acceptance properties; our dispositions to accept the instances of the T-schema, in the case of truth) that make our words mean what they mean, and that happens because the instances of the T-schema needed to know whether p is true or not are not present in our truth theory; Horwich claims that, then, it is conceptually impossible to know whether p is true.

In sum, since we do not accept the paradoxical instances of the T-schema and, hence, the minimalist theory of truth does not contain them as axioms, it is conceptually impossible to know any fact concerning the truth value of the paradoxical propositions; its truth value is indeterminate in Horwich's sense.

At this point, though, the contingent Liar seems to be a problem for Horwich's proposal. As I just said, we are not disposed to accept some of the instances of the T-schema because we realize that they lead to inconsistency.

According to the sometimes called *Simple Conditional Analysis*[7], a subject S has the disposition to accept a given instance of the T-schema when confronted with it iff S would accept it if it were the case that S was confronted with it. So, to say that we do not have the disposition to accept the paradoxical instances of the T-schema is to say that we would not accept such instances in case we were confronted with them. That may be so in the case of The Liar; we may claim that when a subject S is in front of the instance of the T-schema applied to The Liar, S can conclude *a priori* that such instance is not acceptable, because it leads to a contradiction.

As we saw in chapter 1, some paradoxical sentences are not intrinsically paradoxical; that means that they are paradoxical depending on some features of the world; take, for example, the sentence

(3) the sentence written on the blackboard of room 202 is not true.

If the world is such that a token of this sentence is written on the blackboard of room 202, then the sentence is paradoxical but if, for example, the sentence

[7]Defended, for example, in Quine (1986).

written on the blackboard of room 202 is '2+2=4', then (3) is just false. Now, when a subject S is confronted with the instance of the T-schema applied to (3), she cannot say *a priori* if this instance is acceptable or not.

Horwich could try to overcome this difficulty by idealizing the situation; we could idealize the subject S and suppose that she has access to all the relevant facts (in our example, she would have access to room 202). This, though, will not do due to the notion of indeterminacy that Horwich has in mind; it is not just that we are not able to know the semantic value of the Liar sentence (or the ascriptions of vague predicates to borderline cases), but it is conceptually impossible to know it. Let's elaborate this point.

As Field (2010b) has noticed, Horwich's proposal implies that "the concept of an omniscient being is conceptually incoherent" (Field 2010b, p. 2). For an omniscient being would know the location of the boundary between the objects that satisfy a given vague predicate and the objects that do not but, since it is conceptually impossible to know such boundary, we must conclude that the notion of an omniscient being cannot be conceptually coherent. Indeed, Horwich claims that an omniscient being (and any being with a different language of ours) with a language L can only judge whether a given term α (of our language) is true of a given object k via a term β of L with the same meaning as α. That means, having in mind Horwich's theory of meaning, that α and β must have the same "conceptual role" (Horwich 2005a, p. 96). But, then, α and β will be governed by the same fundamental facts about use and, therefore, if α is a vague predicate, then the speaker of L will not be able to find the location of its sharp boundaries, for he will be neither capable of finding the location of the sharp boundaries of β.[8]

The same story applies to the truth predicate; since the instances of the T-schema are the fundamental facts about the use of the truth predicate, any being capable of knowing the truth value of the Liar would have to be using a truth predicate governed by a theory containing the Liar instance of the T-schema, but then this being would not be using our truth predicate, and what this being would have knowledge of would not be the truth value of the Liar sentence, but something else.

[8]This account, as Field (2010b) claims, has at least two problems: first, it is not clear how we can acquire new terms from terms that are untranslatable to our language and, second, even if I have not found yet a synonym of a given expression, it seems clear that I am perfectly capable of pointing to some objects to which it does not apply; for instance, following Field's example, if I hear a bunch of mathematicians employing a word whose meaning is unknown to me I do not need to find a synonym in my own idiolect to legitimately believe that the word does not apply to, say, snails.

Returning now to the problem that the contingent Liar poses to Minimalism, recall that Horwich could try to idealize the situation and suppose that when we evaluate whether a given instance of the T-schema is in our theory of truth we know all the relevant facts. We can create now a contingent Liar whose paradoxicalitiy depends on an indeterminate sentence in Horwich's sense. Take, thus, the proposition expressed by the following sentence:

(L) A and L is not true

where A is any sentence whose truth value is essentially unknowable to us; A can be the Liar sentence itself, or any paradoxical sentence, or an ascription of a vague predicate to a borderline case. Now, since L is paradoxical (that is, its instance of the T-schema can be used to obtain a contradiction via liar-like arguments) just in case A is true and it is conceptually impossible to know the truth value of A, it is conceptually impossible to know whether the L instance of the T-schema should be in the minimalist theory of truth. Hence, even idealizing the situation and supposing that the subject S who is taking the decision whether to incorporate or not a given instance of the T-schema into her theory of truth has all the possible information available in front of her, even in a case like that there will be undecided cases like L.

As we will see in section 6.4, Horwich has abandoned the idea of determining which instances of the T-schema are in the minimalist theory of truth by using dispositional ideas and has tried, instead, to use the notion of grounding.

6.3 The Generalization Problem

6.3 The Generalization Problem and Minimalism

I want to have a look, now, at one of the main problems of Minimalism, which is related to the Liar paradox; the Generalization Problem. As Gupta (1993a,b), Soames (1997, 1999), Armour-Garb (2004, 2010) and Raatikainen (2005), among others, have noted,[9] Minimalism is too weak and has serious problems for explaining many generalizations about truth. That is a major difficulty for minimalists, for it means that the instances of the T-schema are no longer explanatorily fundamental. Consider, for example, the following claim:

(ID) Every proposition of the form $\alpha \rightarrow \alpha$ is true.

[9] A version of the same problem was put forward by Tarski (1983, p. 257).

Can (ID) be derived from all the instances of the T-schema, that is, from the minimalist theory of truth? Certainly, we can derive each instance of (ID), for every proposition p, but that does not mean that we can derive the general fact, expressed by (ID), that all propositions imply themselves. As Scott Soames puts it:

> Because the minimal theory is just a collection of instances, it is conceivable that one could know every proposition in the theory and still be unable to infer [(ID)] because one is ignorant about whether the propositions covered by one's instances are all the (relevant) propositions there are. For example, given only the minimal theory, one might think: perhaps there are more propositions and the [truth predicate] applies differently to them. A person in such a position has no guarantee of [(ID)] and might lack sufficient justification for accepting it. (Soames 1999, p. 247)

The idea is nicely captured in proof-theory. Let me sketch it. Suppose you have a truth theory, $\mathcal{T}$, that consists of all the instances of the T-schema applied to a certain language $\mathcal{L}$ which, for simplicity, we can suppose does not contain the truth predicate. Let $\mathcal{N}$ be a model for $\mathcal{L}$ that fixes the truth values of the sentences without the truth predicate. Suppose, now, in search of a contradiction, that in $\mathcal{T}$ we have a proof, say $\mathcal{P}$, that every sentence of the form $\alpha \rightarrow \alpha$ is true. $\mathcal{P}$, since it is a proof, will be finite and, hence, it will contain a finite number n of instances of the T-schema, say:

1. $< \phi_1 >$ is true iff ϕ_1,
2. $< \phi_2 >$ is true iff ϕ_2,
3. $\cdots$

   $\vdots$

n. $< \phi_n >$ is true iff ϕ_n.

Consider, now, the following set:

$$\Phi = \{\psi : \psi \text{ is true in } \mathcal{N} \text{ and either } \psi \text{ or } \neg\psi \text{ is among } \{\phi_1, \phi_2, \cdots, \phi_n\}\}$$

Notice that, then, if we extend $\mathcal{N}$ by adding Φ as the interpretation of the truth predicate for $\mathcal{L}$, all the instances of the T-schema $1, 2, \cdots, n$ above will be true in the extended model. Let $\mathcal{M}$ be such an extended model. This means that, for any $\chi \in \mathcal{L}$ such that $\chi \rightarrow \chi \notin \{\phi_1, \phi_2, \cdots, \phi_n\}$, $Tr\ulcorner\chi \rightarrow \chi\urcorner$ will not be true in $\mathcal{M}$. That is, we just constructed a model, $\mathcal{M}$, that makes

true all the sentences in $\mathcal{P}$, and hence makes also true that every sentence of the form $\alpha \rightarrow \alpha$ is true but that, on the other hand, has an interpretation of the truth predicate such that, for some sentence χ, $\chi \rightarrow \chi$ is not in this interpretation. Since this is a contradiction, we can conclude that we cannot have a proof that shows that every sentence of the form $\alpha \rightarrow \alpha$ is true.[10]

That means that the minimalist theory of truth is not enough to explain all our uses of the truth predicate, because it cannot explain our acceptance of (ID) only in terms of the instances of the T-schema and some basic logical principles (not involving the truth predicate). Even more, given Horwich's concept of indeterminacy, it is indeterminate whether (ID); thus, it is conceptually impossible to know that (ID), for the impossibility of being justified in believing that (ID) has its roots in the fundamental facts which fix the meaning of 'truth'. As another specific consequence of this, notice that we can only derive each instance of the Principle of Bivalence and not the general principle itself.

In a postscript to his (1998), Horwich makes a first attempt to solve this question:

> [I]t is plausible to suppose that there is a truth-preserving rule of inference that will take us from a set of premises attributing to each proposition some property, *F*, to the conclusion that all propositions have *F*. No doubt this rule is not logically valid, for its reliability hinges [...] on the nature of propositions. But it is a principle we do find plausible. (Horwich 1998b, p. 137)

This explanation remains rather mysterious. First, nothing is said about which feature of propositions is responsible of the plausibility of such a rule.[11] And, second, as Raatikainen (2005) claims, there are several problems that cannot be easily overcome. In the passage above Horwich seems to have in mind some version of the ω-rule, a rule of inference which allows us to deduce a general conclusion concerning some domain of objects from an infinite set of premises ascribing to each object of the domain some given property. One of the features of this kind of rule is that it has an infinitary nature, so it can hardly be used by a human being. Thus, it cannot explain our acceptance of general claims about truth. On the other hand, as Horwich himself admits in his (1998, fn. 4 in page 20), the minimal theory of truth is not a set,[12] for

[10] For details see, for example, Horsten (2011, p. 69) or Halbach (2011, p. 75).

[11] Moreover, this might seem to be at odds with Horwich neutral position with respect to the nature of propositions (see Horwich 1998b, pp. 16–17).

[12] The argument is as follows: suppose the minimal theory is a set. Notice that for every subset X of it, we can define a proposition (call it the *characteristic*

it is too large to be a set, which certainly implies that it is uncountable. But rules of reasoning like the ω-rule require that every element of the universe be named, which is impossible if the intended universe is uncountable.[13]

Horwich (2010c, pp. 43-45, 92-96) follows a new but related strategy to address this problem; instead of looking for a rule of inference, Horwich proposes a further explanatory premise that, on the one hand, allows us to explain our acceptance of general facts concerning truth and, on the other hand, does not involve the truth predicate, for that would jeopardize the minimalist character of Horwich's theory of truth. It should be stressed that this is not an unfamiliar point, for we need principles concerning other phenomena to explain all facts about truth; what is important, though, is that such principles do not use the truth predicate, for if they did, they would show that we need to go beyond the instances of the T-schema in order to explain all facts concerning truth. Horwich proposes the following extra premise:

(P1) Whenever we are disposed to accept, for any proposition of structural type F (henceforth *F-proposition*), that it is G (and to do so for uniform reasons) then we will be disposed to accept that every F-proposition is G.

Furthermore, this premise is restricted to structural kinds of propositions F and properties G that satisfy the following condition:

(C) We cannot conceive of there being additional Fs —beyond those Fs we are disposed to believe are G— which we would not have the same sort of reason to believe are Gs.

proposition of X) which is true if, and only if, all propositions of X are true. Since, if two subsets of the theory are different, so are their characteristic propositions, we can define a 1 to 1 function from the power set of the minimal theory to a subset of the minimal theory (the one that assigns to each subset of the theory the instance of the T-schema applied to its characteristic proposition). That means, intuitively, that the power set of the minimal theory is at most as large as the minimal theory, which, as Cantor's Theorem shows, is not possible. Hence, we conclude, the minimal theory is not a set.

[13] I am setting aside, here, idealizations where, even when the universe in uncountable, we take every object to have itself as a name. In a sense, in these situations we would have an uncountable language, but, as I said, this is an idealization that ignores the fact that the language is to be used by human beings. This last remark, though, is very relevant for the discussion here and, hence, this idealization cannot be hold in this context.

Now, to see how this works according to Horwich, interpret F as $\alpha \rightarrow \alpha$ and G as truth. First notice that they satisfy the requisite (C), for the rules that account for our belief that all propositions of the form $\alpha \rightarrow \alpha$ are true are uniform and do not depend on the proposition p. Furthermore, granted that we are disposed to accept that every proposition of the form $\alpha \rightarrow \alpha$ is true, (P1) allows us to infer that we are disposed to accept that all such propositions are true and, hence, to explain why we accept that all such propositions are true. Finally, since (P1) does not involve the truth predicate, the explicative fundamentality of the minimal theory is preserved.

Armour-Garb (2010) criticizes and rejects this solution to the Generalization Problem. He claims that the extra explanatory premiss should mention the awareness of the fact that we are disposed to accept, for every F-proposition, that it is G:

(P2) Whenever we are disposed to accept, for any F-proposition, that it is G (and to do so for uniform reasons) and we are aware of this fact (that is, we are aware that we are disposed to accept, for any F-proposition, that it is G), then we will be disposed to accept that every F-proposition is G.

The reason for that is that being disposed to accept a given collection of facts is consistent with not knowing the existence of such a disposition and, hence, someone who is disposed to accept, for any F-proposition, that it is G, will accept that all F-propositions are G only if she knows that she has the disposition to accept, for any F-proposition, that it is G. Let's focus now on the partial instance of (P2) where G is interpreted for 'true':

(P3) Whenever we are disposed to accept, for any F-proposition, that it is true (and to do so for uniform reasons) and we are aware of this fact (that is, we are aware that we are disposed to accept, for any F-proposition, that it is true), then we will be disposed to accept that every F-proposition is true.

Armour-Garb claims that, then, we need to clarify in what consists being aware of the fact that we are disposed to accept, for any F-proposition, that it is true. And he proposes the following:

> For one to be aware of the fact that, for every F-proposition, she is disposed to accept that it is true is for that person to be aware of the fact that she is disposed to accept that every F-proposition is true. (Armour-Garb 2010, p. 700)

Thus, (P3) becomes:

(P4) Whenever we are disposed to accept, for any *F*-proposition, that it is true (and to do so for uniform reasons) and we are aware of the fact that we are disposed to accept that every *F*-proposition is true, then we will be disposed to accept that every *F*-proposition is true.

The problem with (P4) is that it is circular; it infers that we have a certain disposition from the fact that we are aware that we have such a disposition.

I agree with Armour-Garb that we need to be aware of the fact that we have the relevant disposition in order to be able to derive the desired generalizations. But, as far as I can see, it is in the spirit of condition (C) to guarantee this awareness; if we become to be convinced by (C) it is because *we tried* to conceive some *F*s not being *G* and it is in this process of trying that we become aware that, in front of every *F* we would be disposed to accept that it is *G*.

Notice that, then, if we accept Armour-Garb's analysis of *being aware*, then (C) alone already does all the job. For Armour-Garb is claiming that

1. being aware of the fact that we are disposed to accept, for every *F*-proposition, that it is true,

is the same as

2. being aware of the fact that we are disposed to accept that every *F*-proposition is true.

This is the reason why, according to Armour-Garb, Horwich's premiss (P) is eventually circular. But if this analysis is right (that is, 1=2), we do not need (P) at all, we just need condition (C) which, as I said, implies 1 and, hence, according to Armour-Garb, implies 2, which is what Horwich needs to be able to explain our acceptance of generalizations about truth (at least if we concede that being aware of *A* implies *A*).

There is another way out of the Generalization Problem. Field (2001, 2006)[14] proposes to understand schemas as something more than the totality of its instances:

> Typically when we advocate a schema [...] we are not merely advocating the collection of instances that happen to be instantiated in our language, we are expressing a commitment to continue to accept new instances as we expand the language. (Field 2006, p. 12)

[14]The idea we are going to see was first introduced, though used in the context of number theory and set theory, in Feferman (1991).

The idea, then, is to introduce schematic letters to the language and use these schematic letters in our reasoning. Then, we need, first, a rule of substitution that allows us to substitute particular sentences for schematic letters in schemas. And, finally, we need an inference rule that allows us to infer generalizations from schemas (following certain restrictions). Thus, for example, we can infer 'a disjunction is true if, and only if, at least one of its disjuncts are true' from "p or q' is true if, and only if, 'p' is true or 'q' is true' (see Field 2006 for more details).

This strategy is not available to Horwich as long as he regards the instances of the T-schema and not the schema itself as epistemologically, explanatorily and conceptually fundamental. It is not clear wether Minimalism can adopt Field's proposal without suffering major changes. Horwich, in his (2010, fn. 15 in p. 95) mentions reasoning with schemas but sticks to his solution in order to maintain the fundamental role of the instances of the T-schema. Anyway, the discussion above suggests that Minimalism can face the generalization problem, at least in the way hitherto presented, with reasonable expectations of success.

Unfortunately, though, recall that Horwich faces the Liar paradox with the restriction of the minimalist theory; then, since there will be instances of the T-schema that will not be in the theory, I will not be able to make, in principle, any generalizations about truth following the previous strategy. Besides, I will not be able to use the truth predicate to express agreement with whatever you said if you said some of the sentences whose instances of the T-schema are not in the minimalist theory. So, eventually, the truth predicate is impaired beyond any apparent hope. As far as I can see, the only strategy available to Horwich is the one he already suggested in his (1998); Horwich can claim that cases like the Liar sentence are "few and far between; so the utility of truth as a device of generalization is not substantially impaired by their existence" (Horwich (1998b, p. 42)). Moreover, we are in front of one of the hardest paradoxes in philosophy of logic and no happy solution should be expected, we know that we will have to give up some things which we are not willing to give up to. Still, it would be much better if we did not have to give up something that undermines the very reason, according to Horwich, of having the truth predicate in our languages and, hence, that undermines one of the main arguments posed by deflationists points of view.

6.3 Minimalism and the Liar

As I said in section 6.2, Horwich's strategy in front of the Liar consists of restricting the instances of the T-schema that constitute the minimalist theory of truth so that no paradox can be formulated; what I called before *the*

paradoxical instances of the T-schema must be ruled out of the truth theory. Then, though, a natural question arises: which instances of the T-schema are to count as paradoxical? Horwich (1998b, p. 42) proposes two conditions that this restriction should meet:

Maximality Instances of the T-schema cannot be excluded unnecessarily; the minimal theory of truth should be, if possible, a maximal consistent collection of instances of the T-schema.

Specification There must be a constructive specification of the instances of the T-schema excluded from the minimal theory of truth. Such specification should be as simple as possible.

Unfortunately, though, McGee (1992) showed that Maximality is not enough to determine which instances of the T-schema should be included in the minimalist theory for, given any consistent set Δ of sentences, there is a maximal consistent set Γ of instances of the T-schema which entails each one of the sentences in Δ.

Let me sketch the proof. First, we use the Diagonalization Lemma to find, for each sentence δ in Δ, a sentence B_δ such that $B_\delta \leftrightarrow (\delta \leftrightarrow Tr(\ulcorner B_\delta \urcorner))$, which implies, by propositional logic, $\delta \leftrightarrow (B_\delta \leftrightarrow Tr(\ulcorner B_\delta \urcorner))$. The idea is that we can find an instance of the T-schema materially equivalent to each one of the members of a given set of sentences.[15] Now, take all the consistent sets of instances of the T-schema that include all the biconditionals of the form $B_\delta \leftrightarrow Tr(\ulcorner B_\delta \urcorner)$ for each δ in Δ. They are partially ordered by the inclusion relation and each chain of this order has a least upper bound (just take the union of all the sets that form the chain). Thus, applying Zorn's Lemma we conclude that this ordered set has a maximal Γ, which is the set we were looking for.

Hence, given two incompatible consistent sets of sentences Δ_1 and Δ_2 we can find two maximal consistent sets of instances of the T-schema Γ_1 and Γ_2 such that the former entails every member of Δ_1 and the latter entails every member of Δ_2. Hence, if we are just looking for maximal consistent sets of instances of the T-schema we have no way to choose between Γ_1 and

[15] Notice that the idea McGee uses is similar to the one used in Curry's paradox. In this paradox we use $\gamma \leftrightarrow (Tr\ulcorner\gamma\urcorner \rightarrow \delta)$, where δ is just any sentence, in order to conclude δ. The fact that we can find an instance of the T-schema necessarily equivalent to any sentence leaves us equally uncomfortable for, if we think that any instance of the T-schema is analytic, that means that we can find, for any sentence whatsoever of our language, an analytic (and hence, necessary) sentence necessarily equivalent to it.

Γ_2, although they are incompatible.[16] What that means is that "the mere desire to preserve as many instances of [the T-schema] as possible will give us too little to go on in constructing" (McGee 1992, p. 237) the minimalist theory of truth. Hence, we are led to the second condition above: we need to be able to specify a particular maximal consistent set of instances of the T-schema.

As Gauker (1999) claims, we can easily see a particular case of McGee's result that might be intuitively easier to grasp. Consider these two sentences:

(λ_1) λ_2 is true.

(λ_2) λ_1 is not true.

They can easily be showed to be paradoxical as follows:

1.	λ_1 is true	Supposition
2.	'λ_2 is true' is true	Identity
3.	λ_2 is true	λ_1-instance of the T-schema
4.	'λ_1 is not true' is true	Identity
5.	λ_1 is not true	λ_2-instance of the T-schema
6.	λ_1 is not true	Reductio 1 and 5
7.	'λ_1 is not true' is true	λ_2-instance of the T-schema
8.	λ_2 is true	Identity
9.	'λ_2 is true' is true	λ_1-instance of the T-schema
10.	λ_1 is true	Identity
11.	Contradiction	6 and 9

The instances of the T-schema we used in this argument are the λ_1-instance (steps 3 and 9) and the λ_2-instance (steps 5 and 7). Clearly it is enough to remove one of them from the theory of truth in order to avoid the paradox generated by λ_1 and λ_2. But the Maximality principle above does not tell us which one we are supposed to remove; so, supposing for a moment that λ_1 and λ_2 are the only sentences in our language that can generate a paradox, we would have two equally good maximal consistent sets of instances of the T-schema and no way of deciding which one is our truth theory.

Future contingents and Curry's paradox give us another example of a situation in which maximality alone fails to determine a unique set of instances of the T-schema. Take the sentences

(s_1) Tomorrow there will be a sea battle

[16] See McGee (1992) and Weir (1996) for more details on McGee's proof.

(s_2) Tomorrow there will not be a sea battle

which are incompatible in the sense that they cannot be true at the same time. Now consider Curry's paradox for each one of them; with the aid of Curry's reasoning we will be able to conclude (s_1) and (s_2), which would allow us to conclude a contradiction. But one of (s_1) or (s_2) will be the case and hence, it could be argued, it would be safe to conclude it. This means that we should remove the instance of the T-schema of the (s_i) that will not be the case; clearly, then, maximality is not enough to determine which instance is to be kept.

6.4 Horwich's Proposal

6.4 The Construction

Horwich has tried to overcome the difficulties posed by the Liar paradox to his theory of truth by offering, in his (2010a), a construction which, although not being maximal, would follow a constructive specification, which was the other condition, apart from maximality, that Minimalism was supposed to follow when restricting the T-schema in front of the Liar. Let's quote the full text:

> We might say that our language L is the limit of the expanding sub-languages L_0, L_1, L_2,... where L_0 lacks the truth predicate; L_1 (which contains L_0) applies it, via the equivalence schema, to the grounded propositions of L_0; similarly, L_2 applies it to the grounded propositions of L_1; L_3 applies it to the grounded propositions of L_2; and so on. Thus an instance of the equivalence schema will be acceptable, even if it governs a proposition concerning truth (e.g. ¡What John said is true¿), as long as the proposition is grounded.
>
> But which propositions of L_0, L_1, L_2, etc. are the grounded ones? They are those that are rooted, as follows, in the *non*-truth-theoretic *facts*. Within L_0, a proposition is grounded just in case the non-truth-theoretic facts either entail that proposition or entail its negation; thus *all* the propositions of L_0 are grounded. Within L_1, a proposition is grounded just in case it, or its negation, is entailed by a combination of those L_0-grounded facts and the (truth-theoretic) facts of L_1 that are 'immediately' entailed by them via the just legitimised instances of the equivalence schema (which are its applications to the grounded propositions of L_0). Similarly, within L_2, a proposition is grounded just in case it, or its negation, is entailed by a combination of those L_1-grounded facts and the facts of L_2 that are 'immediately' entailed

> by them via the just legitimised instances of the equivalence schema (which are its applications to the grounded propositions of L_1). And so on. (Horwich 2010a, p. 90)

Horwich, thus, describes a construction similar to the one proposed in Kripke (1975) and takes the grounded sentences to be the ones whose instances of the T-schema constitute the minimalist theory of truth. This already raises some doubts about whether a deflationist can use the notion of groundedness in order to specify its theory of truth. For the moment though, let's think of this construction as a mere technicality. On the other hand, notice that the quote leaves the distinction between sentences and propositions (or even facts) rather confused. For convenience I will speak about the sentences in the languages, not the propositions.

Let us see, then, if we can have a deeper look at this construction. For perspicuity, let us suppose we have a classical first-order language $\mathcal{L}$, the base language, and an expanded language $\mathcal{L}^+ = \mathcal{L} \cup \{Tr\}$ with a truth predicate Tr and suppose, furthermore, that for every formula $\phi \in \mathcal{L}^+$ we can express its canonical name $\ulcorner \phi \urcorner$ in $\mathcal{L}$ via some codification. I will suppose that $\mathcal{L}$ is strong enough to prove the Diagonal Lemma.

Given a model for the base language, $\mathcal{N}$, with domain D, I will use $\langle \mathcal{N}, A \rangle$ to refer to the model of the expanded language $\mathcal{L}^+$ whose interpretation of Tr is A, which will be a set of (codes of) sentences of $\mathcal{L}^+$. I will use $|\alpha|_{\mathcal{M}} = 1$ to mean that the formula α has semantic value 1 in the model $\mathcal{M}$ (and the same for having semantic value 0). Given a set of formulas Γ, I will use $|\Gamma|_{\mathcal{M}} = 1$ to mean that, for every $\gamma \in \Gamma$, $|\gamma|_{\mathcal{M}} = 1$. $\overline{D}$ will be the set of (codes of) sentences of $\mathcal{L}^+$; as I said I am supposing that, via some suitable codification, $\overline{D} \subseteq D$.

Let's begin with the construction. It will consist of a series H_σ of sets of sentences of $\mathcal{L}^+$ defined for every ordinal σ and relative to a model $\mathcal{N}$ for the base language. We need, first, the following definitions.

Definition Let's define the following.

For any set A of formulas of $\mathcal{L}^+$, $A^- = \{\phi \in \mathcal{L}^+ : \neg\phi \in A\}$.

For any $\phi \in \mathcal{L}^+$, T_ϕ is the ϕ-instance of the T-schema, i.e. $Tr\ulcorner\phi\urcorner \leftrightarrow \phi$.

For any set A of sentences of $\mathcal{L}^+$, $T_A = \{T_\phi : \phi \in A \text{ or } \phi \in A^-\}$.

Now, Horwich presents a construction involving a single truth predicate and a series of sets of sentences of $\mathcal{L}^+$ (which he calls *languages*) which we could try to characterize in the following way, given a model $\mathcal{N}$ for the base language and for any ordinal σ,

$$H_0 = \{\phi \in \mathcal{L} : |\phi|_{\mathcal{N}} = 1\}$$
$$H_{\sigma+1} = \{\phi \in \mathcal{L}^+ : H_\sigma \cup T_{H_\sigma} \models \phi\}$$
$$H_\lambda = \bigcup_{\alpha<\lambda} H_\alpha$$

where λ is a limit ordinal.

Horwich claims, in the quote above, that "our language L is the limit of the expanding sub-languages"; he means with that that the formulas in the alleged limit of the sequence are the formulas whose instances of the T-schema constitute the minimalist theory of truth. It is a good guess to suppose that what Horwich has in mind is something similar to what Kripke (1975) presents in his construction; that is, a fixed point of the construction. Hence, we are looking for an ordinal τ such that $H_\tau = H_{\tau+1}$.

If we want to show the existence of a fixed point, it is sufficient to prove that the series is monotone.

Lemma 6.4.1 (Monotonicity) *If $\tau \leq \rho$, then $H_\tau \subseteq H_\rho$.*

Proof The base case with $\rho = 0$ is trivial. If ρ is a limit ordinal, it follows from the definition of the series.

Suppose now that ρ is a successor ordinal, $\sigma + 1$. Let's suppose, as induction hypothesis, that for all ordinals $\theta \leq \sigma$, $H_\theta \subseteq H_\sigma$. We need to show that for all ordinals $\theta \leq \sigma + 1$, $H_\theta \subseteq H_{\sigma+1}$.

Take, thus, $\theta \leq \sigma + 1$. If $\theta = \sigma + 1$ the result follows trivially. If $\theta < \sigma + 1$ then $\theta \leq \sigma$ which, by induction hypothesis, implies that $H_\theta \subseteq H_\sigma$. Suppose, thus, that $\phi \in H_\theta$, then, by the last remark, $\phi \in H_\sigma$, which implies that $H_\sigma \cup T_{H_\sigma} \models \phi$ and, hence, $\phi \in H_{\sigma+1}$. □

Thus, in each H_σ you keep the sentences present in the previous elements of the series and, in any case, you add new formulas with the use of a set of instances of the T-schema.

Theorem 6.4.2 (Fixed point) *There is an ordinal τ such that $H_\tau = H_{\tau+1}$.*

Proof Suppose there is no such fixed point. Then, given Lemma 6.4.1, for each ordinal σ there is a formula ϕ_σ in $\mathcal{L}^+$ such that $\phi_\sigma \notin H_\sigma$ but $\phi_\sigma \in H_{\sigma+1}$. Notice that the formulas in $\mathcal{L}^+$ form a set. Consequently, if we take the function that assigns to each formula of $\mathcal{L}^+$ the subscript of the H_ρ where it appears first, this function has as its domain a set (the set of formulas of $\mathcal{L}^+$) and has as its range the proper class of all ordinals which, for set theoretic considerations, cannot be. □

I will call the fixed point of the construction $\boldsymbol{H}$. Thus, Horwich's theory of truth, the Minimalist theory of truth, is $T_{\boldsymbol{H}}$.

In the following section, I will show that that construction is consistent by showing that $\boldsymbol{H}$ is a subset of a consistent set, namely the fixed point of Kripke's construction using the supervaluational scheme and restricting the candidate extensions of the truth predicate to consistent sets.

6.4 Kripke and Supervaluations

I will introduce, now, Kripke's fixed point construction (as in Kripke 1975) using the supervaluational scheme. As before $\mathcal{N}$ will be a model of the base language with domain D, $\langle \mathcal{N}, A\rangle$ refers to the model of the expanded language $\mathcal{L}^+$ whose interpretation of Tr is A, which will be a set of (codes of) sentences of $\mathcal{L}^+$. Again, I will use $|\alpha|_{\mathcal{M}} = 1$ to mean that the formula α has semantic value 1 in the model $\mathcal{M}$ (and the same for having semantic value 0); thus $|\ |$ is a classical valuation. $\overline{D}$ will be the set of (codes of) sentences of $\mathcal{L}^+$; as I said I am supposing that, via some suitable codification, $\overline{D} \subseteq D$.

Let us first define the supervaluational scheme, which is a third valued valuation $|\ |^s$ that will take as semantic values 0, ½ and 1. For any $\phi \in \mathcal{L}^+$, any model $\mathcal{N}$ for the base language $\mathcal{L}$ and any set of (codes of) sentences X, $|\psi|^s_{\langle \mathcal{N}, X\rangle}$ is defined in the following way, :

$|\psi|^s_{\langle \mathcal{N}, X\rangle} = 1$ iff, for every Y, such that $X \subseteq Y \subseteq \overline{D} - X^-$, $|\psi|_{\langle \mathcal{N}, Y\rangle} = 1$;

$|\psi|^s_{\langle \mathcal{N}, X\rangle} = 0$ iff, for every Y, such that $X \subseteq Y \subseteq \overline{D} - X^-$, $|\psi|_{\langle \mathcal{N}, Y\rangle} = 0$;

$|\psi|^s_{\langle \mathcal{N}, X\rangle} = 1/2$ otherwise.

We can define next the following series of sentences of $\mathcal{L}^+$, for any ordinal σ,

$$\begin{aligned} VF_0 &= \emptyset \\ VF_{\sigma+1} &= \{\phi \in \mathcal{L}^+ : |\phi|^s_{\langle \mathcal{N}, VF_\sigma\rangle} = 1\} \\ VF_\lambda &= \bigcup_{\alpha<\lambda} VF_\alpha \end{aligned}$$

where λ is a limit ordinal.

As in the case of the previous section, we need to show, first, that the construction in monotonic.

Lemma 6.4.3 (Monotonicity, Kripke 1975) *If $\theta \leq \rho$, then $VF_\theta \subseteq VF_\rho$.*

Proof The proof proceeds by induction on ρ. The base case with $\rho = 0$ is trivial. If ρ is a limit ordinal, it follows from the definition of the series.

Suppose, thus, that ρ is a successor ordinal, $\sigma + 1$. Let's suppose, as induction hypothesis, that for all ordinals $\theta \leq \sigma$, $VF_\theta \subseteq VF_\sigma$. We need to show that for all ordinals $\theta \leq \sigma + 1$, $H_\theta \subseteq VF_{\sigma+1}$.

Take, thus, $\theta \leq \sigma+1$. If $\theta = \sigma+1$ the result follows trivially. If $\theta < \sigma+1$ then $\theta \leq \sigma$ which, by induction hypothesis, implies that $VF_\theta \subseteq VF_\sigma$. Thus, it remains to show that $VF_\sigma \subseteq VF_{\sigma+1}$.

Suppose, thus, that $\phi \notin VF_{\sigma+1}$. Then $|\phi|^s_{\langle \mathcal{N}, VF_\sigma\rangle} \neq 1$ and, hence, there is a Y such that $VF_\sigma \subseteq Y \subseteq \overline{D} - VF^-_\sigma$ and $|\phi|_{\langle \mathcal{N}, Y\rangle} \neq 1$.

Now, σ is either 0, a successor ordinal or a limit ordinal. If $\sigma = 0$, $\phi \notin VF_\sigma$ by definition of VF_0.

If $\sigma = \gamma + 1$ then, by induction hypothesis, $VF_\gamma \subseteq VF_\sigma$ and, hence, $VF^-_\gamma \subseteq VF^-_\sigma$ so that $\overline{D} - VF^-_\sigma \subseteq \overline{D} - VF^-_\gamma$. Henceforth, $VF_\gamma \subseteq Y \subseteq \overline{D} - VF^-_\gamma$. Since, by supposition of Y, $|\phi|_{\langle \mathcal{N}, Y\rangle} \neq 1$, this implies that $|\phi|^s_{\langle \mathcal{N}, VF_\gamma\rangle} \neq 1$ and $\phi \notin VF_\sigma$.

Finally, suppose σ is a limit ordinal, λ. Then, by definition of VF_λ, for all $\zeta < \lambda$, $VF_\zeta \subseteq VF_\lambda$ and, hence, $VF^-_\zeta \subseteq VF^-_\lambda$, which in turn implies that $\overline{D} - VF^-_\lambda \subseteq \overline{D} - VF^-_\zeta$. This means that for all $\zeta < \lambda$, $VF_\zeta \subseteq Y \subseteq \overline{D} - VF^-_\zeta$. Since, by supposition, $|\phi|_{\langle \mathcal{N}, Y\rangle} \neq 1$, we conclude that $|\phi|^s_{\langle \mathcal{N}, VF_\zeta\rangle} \neq 1$, which, in turn, implies that $\phi \notin VF_{\zeta+1}$. Finally, since $\zeta + 1 < \lambda$, we conclude that $\phi \notin VF_\lambda$. □

For the same considerations as in Theorem 6.4.2 there will exist a fixed point of the construction, that is, an ordinal ρ such that $VF_\rho = VF_{\rho+1}$. I will call this fixed point $\boldsymbol{VF}$.

Following Kripke (1975) and Field (2008) we can now define variations on the supervaluational scheme by imposing a condition Φ on the candidate extensions of the truth predicate.[17] These restrictions will create other fixed points that will be supersets of $\boldsymbol{VF}$. In order to proceed, we define $|\psi|^{\Phi,s}_{\langle \mathcal{N}, X\rangle}$ more generally:

$|\psi|^{\Phi,s}_{\langle \mathcal{N}, X\rangle} = 1$ iff, for every Y, such that $\Phi(Y)$ and $X \subseteq Y \subseteq \overline{D} - X^-$, $|\psi|_{\langle \mathcal{N}, Y\rangle} = 1$;

$|\psi|^{\Phi,s}_{\langle \mathcal{N}, X\rangle} = 0$ iff, for every Y, such that $\Phi(Y)$ and $X \subseteq Y \subseteq \overline{D} - X^-$, $|\psi|_{\langle \mathcal{N}, Y\rangle} = 0$;

[17] As I said in chapter 3, McGee (1991) also uses these techniques.

$|\psi|^{\Phi,s}_{\langle \mathcal{N},X\rangle} = 1/2$ otherwise

In this definition I am presupposing that there will always be a Y satisfying the condition Φ and such that $X \subseteq Y \subseteq \overline{D} - X^{-}$.[18] Given a condition Φ, I will call VF^{Φ}_{σ} the σ stage of the construction using Φ as the property to be satisfied by the candidate extensions of the Truth predicate. I will call $\boldsymbol{VF}^{\Phi}$ the fixed point of such construction.

We can consider now the following fixed points corresponding to the following conditions:

The vacuous condition: $\boldsymbol{VF}$

Consistency: $\boldsymbol{VF}^{c}$

Closure under classical deduction: $\boldsymbol{VF}^{cd}$

Maximal consistency: $\boldsymbol{VF}^{mc}$

A trivial generalization of Lemma 6.4.3 together with the considerations in Theorem 6.4.2 show that all of $\boldsymbol{VF}$, $\boldsymbol{VF}^{c}$, $\boldsymbol{VF}^{cd}$ and $\boldsymbol{VF}^{mc}$ exist. We must see now that all these fixed points are consistent. The following Lemma offers a sufficient condition on Φ for consistency.

Lemma 6.4.4 (Field 2008, p. 180) *Let λ be the Liar sentence. For any given property Φ, if for every consistent and deductively closed set of sentences Z such that $\lambda \notin Z$ and $\neg\lambda \notin Z$ there are Y_1 and Y_2 such that $\lambda \notin Y_1$, $\lambda \in Y_2$, $\Phi(Y_i)$ and $Z \subseteq Y_i \subseteq \overline{D} - Z^{-}$ $(1 \leq i \leq 2)$, then $\boldsymbol{VF}^{\Phi}$ is consistent.*

Proof Let Φ be any property. Suppose that the antecedent of the Lemma is true.

In order to seek a contradiction suppose, now, that $\boldsymbol{VF}^{\Phi}$ is not consistent. Then there is a smallest ordinal σ such that either $\lambda \in VF^{\Phi}_{\sigma}$ or $\neg\lambda \in VF^{\Phi}_{\sigma}$. Notice that σ must be a successor ordinal, say, $\delta + 1$. This means that VF^{Φ}_{δ} is consistent, closed under classical deduction, $\lambda \notin VF^{\Phi}_{\delta}$ and $\neg\lambda \notin VF^{\Phi}_{\delta}$. The antecedent of the Lemma, then, implies that there are Y_1 and Y_2 such that $\lambda \notin Y_1$, $\lambda \in Y_2$, $\Phi(Y_i)$ and $VF^{\Phi}_{\delta} \subseteq Y_i \subseteq \overline{D} - VF^{\Phi-}_{\delta}$. Therefore,

[18] If it were not the case, the construction would have to be adjusted with a fourth semantic value to represent the situation where there is no appropriate Y. Notice that, if this fourth semantic value is not defined and there are no Y's satisfying the condition Φ and such that $X \subseteq Y \subseteq \overline{D} - X^{-}$, then all sentences would have trivially the values 1 and 0. See Kripke (1975, p. 711) and Field (2008, p. 178) for more details.

$|Tr^{\ulcorner}\lambda^{\urcorner}|^{s}_{\langle \mathcal{N}, VF^{\Phi}_{\delta}\rangle} = 1/2$ and $|\neg Tr^{\ulcorner}\lambda^{\urcorner}|^{s}_{\langle \mathcal{N}, VF^{\Phi}_{\delta}\rangle} = 1/2$. Consequently, $Tr^{\ulcorner}\lambda^{\urcorner} \notin VF^{\Phi}_{\sigma}$ and $\neg Tr^{\ulcorner}\lambda^{\urcorner} \notin VF^{\Phi}_{\sigma}$. But we were supposing that either $\lambda \in VF^{\Phi}_{\sigma}$ or $\neg\lambda \in VF^{\Phi}_{\sigma}$ which, together with the facts that, by the Diagonal Lemma, $\lambda \leftrightarrow \neg Tr^{\ulcorner}\lambda^{\urcorner} \in VF^{\Phi}_{\sigma}$, and that VF^{Φ}_{σ} is deductively closed, implies that either $Tr^{\ulcorner}\lambda^{\urcorner} \in VF^{\Phi}_{\sigma}$ or $\neg Tr^{\ulcorner}\lambda^{\urcorner} \in VF^{\Phi}_{\sigma}$. We have reached, thus, the contradiction we were seeking. □

Corollary 6.4.5 (Field 2008, p. 180)

(i) $\boldsymbol{VF}$ *is consistent.*

(ii) $\boldsymbol{VF}^{c}$ *is consistent.*

(iii) $\boldsymbol{VF}^{cd}$ *is consistent.*

(iv) $\boldsymbol{VF}^{mc}$ *is consistent.*

Proof

(i) Take any set of sentences Z consistent, deductively closed and such that $\lambda \notin Z$ and $\neg\lambda \notin Z$. By Lemma 6.4.4 we need two sets of sentences, Y_1 and Y_2, such that $\lambda \notin Y_1$, $\lambda \in Y_2$ and $Z \subseteq Y_i \subseteq \overline{D} - Z^{-}$. Take Z as Y_1 and the deductive closure of $Z \cup \{\lambda\}$ as Y_2. For (ii) and (iii) we can use the same Y_1 and Y_2, for both are consistent and deductively closed.

(iv) Take a maximal consistent extension of $Z \cup \{\neg\lambda\}$ as Y_1 and a maximal consistent extension of $Z \cup \{\lambda\}$ as Y_2. [19] □

There are several relations that can be established between the fixed points we have presented.

Proposition 6.4.6

(i) $\boldsymbol{VF} \subsetneq \boldsymbol{VF}^{c}$ *(Kripke 1975, p. 711)*

(ii) $\boldsymbol{VF}^{c} \subsetneq \boldsymbol{VF}^{dc}$

(iii) $\boldsymbol{VF}^{dc} \subsetneq \boldsymbol{VF}^{mc}$ *(Kripke 1975, p. 711)*

[19] Notice that this Corollary, together with the previous Lemma (6.4.4), shows that, for the properties considered here (vacuous property, consistency, classical deductive closure and maximal consistency), there will always be a Y satisfying the desired properties for the supervaluational scheme as defined above so that no fourth semantic value is needed.

Proof

(i) $\boldsymbol{VF} \subseteq \boldsymbol{VF}^c$ follows trivially. To see that $\boldsymbol{VF} \neq \boldsymbol{VF}^c$ notice that $\neg(Tr^{\ulcorner}\lambda^{\urcorner} \wedge Tr^{\ulcorner}\neg\lambda^{\urcorner}) \in \boldsymbol{VF}^c - \boldsymbol{VF}$.

(ii) Let us show that $\boldsymbol{VF}^c \subseteq \boldsymbol{VF}^{dc}$ in more detail. We will show that, for each ordinal σ, $VF^c_\sigma \subseteq VF^{dc}_\sigma$ by induction on σ. The 0 and limit cases follow immediately from the definition of the series. For the successor case, suppose that $\sigma = \rho + 1$ and, as induction hypothesis, that $VF^c_\rho \subseteq VF^{dc}_\rho$. We need to show that $VF^c_{\rho+1} \subseteq VF^{dc}_{\rho+1}$. Take $\phi \in VF^c_{\rho+1}$ in order to show that $\phi \in VF^{dc}_{\rho+1}$.

Since $\phi \in VF^c_{\rho+1}$, we have that $|\phi|^{c,s}_{\langle \boldsymbol{\mathcal{N}}, VF^c_\rho\rangle} = 1$, which means that, for all Y, such that Y is consistent and $VF^c_\rho \subseteq Y \subseteq \overline{D} - VF^{c-}_\rho$, $|\phi|_{\langle \boldsymbol{\mathcal{N}}, Y\rangle} = 1$.

Take now any Y closed under classical deduction such that $VF^{dc}_\rho \subseteq Y \subseteq \overline{D} - VF^{dc-}_\rho$. By the induction hypothesis, $VF^c_\rho \subseteq Y \subseteq \overline{D} - VF^{c-}_\rho$. Notice now that, by construction of the series, $VF^{c-}_\rho \neq \emptyset$, which means, together with the fact that Y is closed under classical deduction, that Y is consistent. This implies that $|\phi|_{\langle \boldsymbol{\mathcal{N}}, Y\rangle} = 1$ and, hence, that $|\phi|^{dc,s}_{\langle \boldsymbol{\mathcal{N}}, VF^c_\rho\rangle} = 1$. Therefore, $\phi \in VF^{dc}_{\rho+1}$.

It remains to show that $\boldsymbol{VF}^c \neq \boldsymbol{VF}^{mc}$. Take the sentence $(Tr^{\ulcorner}\lambda^{\urcorner} \rightarrow Tr^{\ulcorner}\lambda \vee \psi^{\urcorner})$, where ψ is any sentence of the ground language such that $|\psi|_{\boldsymbol{\mathcal{N}}} = 0$. By Corollary 6.4.5, $\lambda \notin \boldsymbol{VF}^{dc}$ and, since ψ is a false sentence of the base language, $\lambda \vee \psi \notin \boldsymbol{VF}^{dc}$. This means, by monotonicity, that for each ordinal σ, $\lambda \vee \psi \notin VF^{dc}_\sigma$. Now, if we consider a given Y closed under classical deduction such that $VF^{dc}_\sigma \subseteq Y \subseteq \overline{D} - VF^{dc-}_\sigma$, for some stage σ in the construction of $\boldsymbol{VF}^{dc}$, we have two options. First, if $\lambda \notin Y$, then $|Tr^{\ulcorner}\lambda^{\urcorner}|_{\langle \boldsymbol{\mathcal{N}}, Y\rangle} = 0$ and, hence, $|Tr^{\ulcorner}\lambda^{\urcorner} \rightarrow Tr^{\ulcorner}\lambda \vee \psi^{\urcorner}|_{\langle \boldsymbol{\mathcal{N}}, Y\rangle} = 1$. Second, if $\lambda \in Y$, then, since Y is closed under classical deduction, $\lambda \vee \psi \in Y$ and, hence, $|Tr^{\ulcorner}\lambda^{\urcorner} \rightarrow Tr^{\ulcorner}\lambda \vee \psi^{\urcorner}|_{\langle \boldsymbol{\mathcal{N}}, Y\rangle} = 1$. This implies that $Tr^{\ulcorner}\lambda^{\urcorner} \rightarrow Tr^{\ulcorner}\lambda \vee \psi^{\urcorner} \in \boldsymbol{VF}^{dc}$.

However, there will be consistent candidate extensions of the truth predicate in the construction of $\boldsymbol{VF}^c$ that will contain λ but not $\psi \vee \lambda$, which means that $(Tr^{\ulcorner}\lambda^{\urcorner} \rightarrow Tr^{\ulcorner}\lambda \vee \psi^{\urcorner}) \notin \boldsymbol{VF}^c$.

(iii) $\boldsymbol{VF}^{dc} \subseteq \boldsymbol{VF}^{mc}$ follows from the fact that maximal consistency implies closure under classical deduction. To see that $\boldsymbol{VF}^{dc} \neq \boldsymbol{VF}^{mc}$ notice that $(Tr^{\ulcorner}\lambda^{\urcorner} \vee Tr^{\ulcorner}\neg\lambda^{\urcorner}) \in \boldsymbol{VF}^{mc} - \boldsymbol{VF}^{dc}$. □

Finally, we can now see that $\boldsymbol{H}$ is consistent, as it is just $\boldsymbol{VF}$.

Lemma 6.4.7 *For any ordinal σ, $H_\sigma \subseteq VF_{\sigma+1}$.*

Proof The proof proceeds by induction on σ.

- Base case. If $\sigma = 0$, then if $\phi \in H_0$, by definition of H_0 we have that $\phi \in \mathcal{L}$ and $|\phi|_{\mathcal{N}} = 1$. This means that for every Y such that $VF_0 \subseteq Y \subseteq \overline{D} - VF_0^-$, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$ (just because the value of ϕ, being from the base language, is independent of the choice of the extension of the truth predicate) and, hence, $\phi \in VF_1$.

- Successor case. Suppose now, as induction hypothesis, that $H_\sigma \subseteq VF_{\sigma+1}$. We need to show that $H_{\sigma+1} \subseteq VF_{\sigma+2}$.

  Take $\phi \in H_{\sigma+1}$. This means that $H_\sigma \cup T_{H_\sigma} \models \phi$. What we need to show is that $|\phi|^s_{\langle \mathcal{N}, VF_{\sigma+1}\rangle} = 1$, that is, we need to show that for any Y such that $VF_{\sigma+1} \subseteq Y \subseteq \overline{D} - VF^-_{\sigma+1}$, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$.

  So take any Y satisfying the conditions above. We will show now that $|H_\sigma|_{\langle \mathcal{N}, Y\rangle} = 1$ and that $|T_{H_\sigma}|_{\langle \mathcal{N}, Y\rangle} = 1$. This means that, since we are supposing that $H_\sigma \cup T_{H_\sigma} \models \phi$, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$, which is what we want to prove.

  - $|H_\sigma|_{\langle \mathcal{N}, Y\rangle} = 1$.
    Take $\phi \in H_\sigma$. By induction hypothesis, $\phi \in VF_{\sigma+1}$ and, by monotonicity (Lemma 6.4.3), $\phi \in VF_{\sigma+2}$, which means that $|\phi|^s_{\langle \mathcal{N}, VF_{\sigma+1}\rangle} = 1$ and, hence, for any Y' such that $VF_{\sigma+1} \subseteq Y' \subseteq \overline{D} - VF^-_{\sigma+1}$, $|\phi|_{\langle \mathcal{N}, Y'\rangle} = 1$ so, in particular, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$.
  - $|T_{H_\sigma}|_{\langle \mathcal{N}, Y\rangle} = 1$.
    Take, first, $\phi \in H_\sigma$. We just showed that $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$. But since, by induction hypothesis and construction of Y, $H_\sigma \subseteq VF_{\sigma+1} \subseteq Y$, $\phi \in Y$ and, hence, $|Tr\langle\phi\rangle|_{\langle \mathcal{N}, Y\rangle} = 1$. Consequently, $|Tr^\ulcorner\phi^\urcorner \leftrightarrow \phi|_{\langle \mathcal{N}, Y\rangle} = 1$.
    Take, now, $\phi \in H^-_\sigma$. Then, $\neg\phi \in H_\sigma$, $|\neg\phi|_{\langle \mathcal{N}, Y\rangle} = 1$ and, hence, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 0$. On the other hand, by induction hypothesis, $\neg\phi \in VF_{\sigma+1}$ and, therefore, $\phi \in VF^-_{\sigma+1}$. But, since by construction, $Y \subseteq \overline{D} - VF^-_{\sigma+1}$, $\phi \notin Y$ and, consequently, $|Tr\langle\phi\rangle|_{\langle \mathcal{N}, Y\rangle} = 0$. Therefore, $|Tr^\ulcorner\phi^\urcorner \leftrightarrow \phi|_{\langle \mathcal{N}, Y\rangle} = 1$. □

- Limit case. Suppose now by induction hypothesis that for all ordinals σ, $\sigma < \lambda$, for a given limit ordinal λ, $H_\sigma \subseteq VF_{\sigma+1}$. We need to show that $H_\lambda \subseteq VF_{\lambda+1}$. Take any $\phi \in H_\lambda$. By definition of H_λ, for some ρ, $\rho < \lambda$, $\phi \in H_\rho$. By induction hypothesis it follows that $\phi \in VF_{\rho+1}$ and hence, by definition of VF_λ, $\phi \in VF_\lambda$. This, by monotonicity (Lemma 6.4.3) implies that $\phi \in VF_{\lambda+1}$. □

Lemma 6.4.8 *For any ordinal σ, $VF_\sigma \subseteq H_\sigma$.*

Proof The proof proceeds by induction on σ.

- Base case. Clear from the definition of VF_0.
- Successor case. Suppose now, as induction hypothesis, that $VF_\sigma \subseteq H_\sigma$. We need to show that $VF_{\sigma+1} \subseteq H_{\sigma+1}$.

  Take $\phi \in VF_{\sigma+1}$. This means that $|\phi|^s_{\langle \mathcal{N}, VF_\sigma\rangle} = 1$, that is, for any Y such that $VF_\sigma \subseteq Y \subseteq \overline{D} - VF^-_\sigma$, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$. We need to show that $H_\sigma \cup T_{H_\sigma} \models \phi$.

  So take any model $\langle \mathcal{N}, X\rangle$ such that $|H_\sigma \cup T_{H_\sigma}|_{\langle \mathcal{N}, X\rangle} = 1$. We need to show that $|\phi|_{\langle \mathcal{N}, X\rangle} = 1$. In order to do that it will be enough to see that $VF_\sigma \subseteq X \subseteq \overline{D} - VF^-_\sigma$.

  - $VF_\sigma \subseteq X$.

    Take $\psi \in VF_\sigma$. Then, by induction hypothesis, $\psi \in H_\sigma$, which means that, by supposition, $|\psi|_{\langle \mathcal{N}, X\rangle} = 1$. Moreover, if $\psi \in H_\sigma$, then $Tr^\ulcorner\psi^\urcorner \leftrightarrow \psi \in T_{H_\sigma}$ and, again by supposition, $|Tr^\ulcorner\psi^\urcorner \leftrightarrow \psi|_{\langle \mathcal{N}, X\rangle} = 1$. Therefore $|Tr^\ulcorner\psi^\urcorner|_{\langle \mathcal{N}, X\rangle} = 1$, which implies that $\psi \in X$.
  - $X \subseteq \overline{D} - VF^-_\sigma$.

    Take $\psi \in X$. We want to show that $\psi \notin VF^-_\sigma$. Suppose, in order to seek a contradiction, that $\psi \in VF^-_\sigma$. In that case, $\neg\psi \in VF_\sigma$ and, by induction hypothesis, $\neg\psi \in H_\sigma$ and, hence, $\psi \in H^-_\sigma$. Therefore, by definition of T_{H_σ}, $Tr^\ulcorner\psi^\urcorner \leftrightarrow \psi \in T_{H_\sigma}$ and, hence, by supposition, $|Tr^\ulcorner\psi^\urcorner \leftrightarrow \psi|_{\langle \mathcal{N}, X\rangle} = 1$. But, if $\neg\psi \in H_\sigma$ then, again by supposition, $|\neg\psi|_{\langle \mathcal{N}, X\rangle} = 1$ and, hence, $|\psi|_{\langle \mathcal{N}, X\rangle} = 0$. Consequently $|Tr^\ulcorner\psi^\urcorner|_{\langle \mathcal{N}, X\rangle} = 0$, which means that $\psi \notin X$. Contradiction.

- Limit case. Suppose now by induction hypothesis that for all ordinals σ, $\sigma < \lambda$ (λ a limit ordinal), $VF_\sigma \subseteq H_\sigma$. We need to show that $VF_\lambda \subseteq H_\lambda$.

Take any $\phi \in VF_\lambda$. By definition of VF_λ, for some ρ, $\rho < \lambda$, $\phi \in VF_\rho$. By induction hypothesis it follows that $\phi \in H_\rho$ and hence, by definition of H_λ, $\phi \in H_\lambda$. □

Corollary 6.4.9 $\boldsymbol{H = VF}$

We can see now that interpreting the truth predicate as any extension of $\boldsymbol{H}$ such that is disjoint with $\boldsymbol{H}^-$ provides us with a model for the fixed point.

Lemma 6.4.10 *For all sets of sentences of* $\mathcal{L}^+$, Y, *such that* $\boldsymbol{H} \subseteq Y \subseteq \overline{D} - \boldsymbol{H}^-$, $|\boldsymbol{H}|_{\langle \mathcal{N}, Y\rangle} = 1$.

Proof It follows immediately from the fact that $\boldsymbol{H} = \boldsymbol{VF}$. Suppose $\boldsymbol{VF} = VF_\tau$. Then, $VF_\tau = VF_{\tau+1}$ and, hence, by definition of $VF_{\tau+1}$, every $\phi \in VF_{\tau+1}$ is such that $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$, where $VF_\tau \subseteq Y \subseteq \overline{D} - VF_\tau^-$, which is what we needed to prove. □

Corollary 6.4.11 *For all sets of sentences of* $\mathcal{L}^+$, Y, *such that* $\boldsymbol{H} \subseteq Y \subseteq \overline{D} - \boldsymbol{H}^-$, $|T_{\boldsymbol{H}}|_{\langle \mathcal{N}, Y\rangle} = 1$.

Proof Take, first, $\phi \in \boldsymbol{H}$. By Lemma 6.4.10, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 1$ and, by construction of Y, $\phi \in Y$. This implies that $|Tr^\ulcorner\phi^\urcorner|_{\langle \mathcal{N}, Y\rangle} = 1$ and, hence, $|\phi \leftrightarrow Tr^\ulcorner\phi^\urcorner|_{\langle \mathcal{N}, Y\rangle} = 1$.

Take, now, $\phi \in \boldsymbol{H}^-$. Then, $\neg\phi \in \boldsymbol{H}$ and, by Lemma 6.4.10, $|\phi|_{\langle \mathcal{N}, Y\rangle} = 0$. On the other hand, by construction of Y, $\phi \notin Y$. This implies that $|Tr^\ulcorner\phi^\urcorner|_{\langle \mathcal{N}, Y\rangle} = 0$ and, hence, $|\phi \leftrightarrow Tr^\ulcorner\phi^\urcorner|_{\langle \mathcal{N}, Y\rangle} = 1$. □

We can now prove an important feature of Horwich's fixed point.

Proposition 6.4.12 *For all sentences* ϕ *of* $\mathcal{L}^+$, $\phi \in \boldsymbol{H}$ *if, and only if,* $Tr^\ulcorner\phi^\urcorner \in \boldsymbol{H}$.

Proof Suppose $\boldsymbol{H} = H_\tau$. For the left to right direction, suppose $\phi \in H_\tau$. Then $\phi \leftrightarrow Tr^\ulcorner\phi^\urcorner \in T_{H_\tau}$ and, hence, by *Modus Ponens*, $H_\tau \cup T_{H_\tau} \models Tr^\ulcorner\phi^\urcorner$. That means that $Tr^\ulcorner\phi^\urcorner \in H_{\tau+1}$ and, since H_τ is a fixed point, $Tr^\ulcorner\phi^\urcorner \in H_\tau$.

For the right to left direction, suppose $Tr^\ulcorner\phi^\urcorner \in \boldsymbol{H}$. Then, by Lemma 6.4.10, $|Tr^\ulcorner\phi^\urcorner|_{\langle \mathcal{N}, \boldsymbol{H}\rangle} = 1$, which means that $\phi \in \boldsymbol{H}$. □

Notice also that, as the following Proposition shows, all the axioms of the Minimalist theory are founded, that is, they are in $\boldsymbol{H}$.

Proposition 6.4.13 $T_{\boldsymbol{H}} \subseteq \boldsymbol{H}$

Proof Take any $\phi \leftrightarrow Tr^\ulcorner\phi^\urcorner \in T_{\boldsymbol{H}}$. Clearly, $\boldsymbol{H} \cup T_{\boldsymbol{H}} \models \phi \leftrightarrow Tr^\ulcorner\phi^\urcorner$. Hence, by the fact that $\boldsymbol{H}$ is a fixed point, it follows that $\phi \leftrightarrow Tr^\ulcorner\phi^\urcorner \in \boldsymbol{H}$. □

6.4 The Minimal Theory of Truth

We should ask ourselves now whether the theory of truth Horwich is proposing is satisfactory. Recall that we can now characterize in a precise way which is, according to Horwich (2010a), the minimalist theory of truth: $T_{\boldsymbol{H}}$. This means, consequently, that all that can be said about pure truth should follow from $T_{\boldsymbol{H}}$. Hence, we could take the set $\mathcal{T} = \{\phi : T_{\boldsymbol{H}} \vdash \phi\}$ to be everything that can be said with respect to pure truth. We will see later which is the set that represents what can be known about truth at all.

It is well known that $\boldsymbol{VF}$ has many unsatisfactory properties, which, as we will see in the following lines, are inherited by $\mathcal{T}$. Here there are four laws we might expect to have in $\mathcal{T}$:

1. For any sentence x; $Tr\ulcorner\neg\neg x\urcorner$ if, and only if, $Tr\ulcorner x\urcorner$.
2. For any sentences x, y; $Tr\ulcorner x \vee y\urcorner$ if, and only if, $Tr\ulcorner x\urcorner$ or $Tr\ulcorner y\urcorner$.
3. For any sentences x, y; $Tr\ulcorner x \wedge y\urcorner$ if, and only if, $Tr\ulcorner x\urcorner$ and $Tr\ulcorner y\urcorner$.
4. For any sentences x, y; if $Tr\ulcorner x \rightarrow y\urcorner$ and $Tr\ulcorner x\urcorner$, then $Tr\ulcorner y\urcorner$.

Unfortunately, though, none of these laws are satisfied neither by $\mathcal{T}$ nor by $\boldsymbol{H}$; specifically, the problem is that we can find some instances that are not in it.

Thus, with respect to 1, notice that, by the fact that $\lambda \notin \boldsymbol{H}$ and Corollary 6.4.11, $|T_{\boldsymbol{H}}|_{\langle \mathcal{N}, \boldsymbol{H}\cup\{\lambda\}\rangle} = 1$. But since $|Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner|_{\langle \mathcal{N}, \boldsymbol{H}\cup\{\lambda\}\rangle} = 0$, $T_{\boldsymbol{H}} \not\models Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner$ and, hence, $Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner \notin \mathcal{T}$. Similarly, we can see that $Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner \notin \boldsymbol{H}$. By the fact $\lambda \notin \boldsymbol{H}$, Lemma 6.4.10 and Corollary 6.4.11, $|T_{\boldsymbol{H}} \cup \boldsymbol{H}|_{\langle \mathcal{N}, \boldsymbol{H}\cup\{\lambda\}\rangle} = 1$. Again, since $|Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner|_{\langle \mathcal{N}, \boldsymbol{H}\cup\{\lambda\}\rangle} = 0$, we conclude that $T_{\boldsymbol{H}} \cup \boldsymbol{H} \not\models Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner$, which, since $\boldsymbol{H}$ is a fixed point, yields that $Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner \notin \boldsymbol{H}$.

The following table lists the failure of the laws 1-4, the models that show such failure and the corresponding examples of sentences that are neither in $\boldsymbol{H}$ nor in $\mathcal{T}$ (as before, λ is the Liar sentence).

Law	Model	Sentence
$1^{\rightarrow}$	$\langle \mathcal{N}, \boldsymbol{H} \cup \{\lambda\}\rangle$	$Tr\ulcorner\lambda\urcorner \rightarrow Tr\ulcorner\neg\neg\lambda\urcorner$
$1^{\leftarrow}$	$\langle \mathcal{N}, \boldsymbol{H} \cup \{\neg\neg\lambda\}\rangle$	$Tr\ulcorner\neg\neg\lambda\urcorner \rightarrow Tr\ulcorner\lambda\urcorner$
$2^{\rightarrow}$	$\langle \mathcal{N}, \boldsymbol{H}\rangle$	$Tr\ulcorner\lambda \vee \neg\lambda\urcorner \rightarrow (Tr\ulcorner\lambda\urcorner \vee Tr\ulcorner\neg\lambda\urcorner)$
$2^{\leftarrow}$	$\langle \mathcal{N}, \boldsymbol{H} \cup \{\lambda\}\rangle$	$(Tr\ulcorner\lambda\urcorner \vee Tr\ulcorner\lambda\urcorner) \rightarrow Tr\ulcorner\lambda \vee \lambda\urcorner$
$3^{\rightarrow}$	$\langle \mathcal{N}, \boldsymbol{H} \cup \{\lambda \wedge \lambda\}\rangle$	$Tr\ulcorner\lambda \wedge \lambda\urcorner \rightarrow (Tr\ulcorner\lambda\urcorner \wedge Tr\ulcorner\lambda\urcorner)$
$3^{\leftarrow}$	$\langle \mathcal{N}, \boldsymbol{H} \cup \{\lambda\}\rangle$	$(Tr\ulcorner\lambda\urcorner \wedge Tr\ulcorner\lambda\urcorner) \rightarrow Tr\ulcorner\lambda \wedge \lambda\urcorner$
$4^{\rightarrow}$	$\langle \mathcal{N}, \boldsymbol{H} \cup \{\lambda \rightarrow \neg\lambda, \lambda\}\rangle$	$(Tr\ulcorner\lambda \rightarrow \neg\lambda\urcorner \wedge Tr\ulcorner\lambda\urcorner) \rightarrow Tr\ulcorner\neg\lambda\urcorner$

Recall that in section 6.3 we saw the generalization problem: even without consistency problems (that is, with no restrictions on the T-schema over some language $\mathcal{L}$) the minimalist theory of truth would not be able to prove general statements about truth, but only its instances. We eventually concluded that this problem might be satisfactorily overcome.

But now we can see the difficulties that the Liar poses to the minimalist theory of truth with all its virulence. First, all the principles 1-4 will not be in $\mathcal{T}$, which means, according to Horwich, that they will be principles about truth that will remain unknown to us; as a matter of fact, they will be conceptually impossible to know.

Let us see what does exactly mean to say that laws 1-4 are not in $\boldsymbol{H}$. We have seen that any model $\mathcal{M}$ for $\mathcal{L}^+$ with an extension of the truth predicate, $Tr^{\mathcal{M}}$, is such that $\boldsymbol{H} \subseteq Tr^{\mathcal{M}} \subseteq \overline{D} - \boldsymbol{H}^-$ satisfies $T_{\boldsymbol{H}}$ —which is the minimalist theory of truth— and $\boldsymbol{H}$. We have to ask ourselves, first, who should we understand $\boldsymbol{H}$ is. The way the construction is devised makes it natural to consider the set $\boldsymbol{H} \cup \boldsymbol{H}^-$ as the set of grounded sentences; specifically, $\boldsymbol{H}$ is the set of determinately true sentences (that is, supposing there are not vague predicates nor other sources of indeterminacy in Horwich's epistemic sense, the sentences which are conceptually possible to know) and $\boldsymbol{H}^-$ is the set of determinately false sentences (that is, the sentences whose negations are determinately true). Recall that all of these are relative to a given ground model $\mathcal{N}$. To continue with this picture, notice that, as I said, we can interpret H_0 as the theory of all extra semantic facts given by the ground model $\mathcal{N}$ and define $\boldsymbol{T} = \{\phi : T_{\boldsymbol{H}} \cup H_0 \vdash \phi\}$ as everything that can be known about truth at all. Then, it is natural to expect the following proposition (for any given sets of sentences Γ and Δ, I will use $\Gamma \models \Delta$ to mean that, for every $\delta \in \Delta$, $\Gamma \models \delta$).

Proposition 6.4.14 $\boldsymbol{T} = \boldsymbol{H}$

Proof We will prove, first, that $\boldsymbol{H} \subseteq \boldsymbol{T}$. In order to do that we will show that for every ordinal σ, $T_{\boldsymbol{H}} \cup H_0 \models H_\sigma$. The proof proceeds by induction on σ. The base case and the limit cases are clear.

For the successor case, suppose that $T_{\boldsymbol{H}} \cup H_0 \models H_\sigma$ in order to show that $T_{\boldsymbol{H}} \cup H_0 \models H_{\sigma+1}$. Take $\phi \in H_{\sigma+1}$. By definition of $H_{\sigma+1}$, $H_\sigma \cup T_{H_\sigma} \models \phi$. Now, from the induction hypothesis and the fact that $T_{H_\sigma} \subseteq T_{\boldsymbol{H}}$, it follows that $T_{\boldsymbol{H}} \cup H_0 \models H_\sigma \cup T_{H_\sigma}$. Hence, $T_{\boldsymbol{H}} \cup H_0 \models \phi$.

Let us show, now, that $\boldsymbol{T} \subseteq \boldsymbol{H}$. Suppose $\phi \notin \boldsymbol{H}$. Then, for no ρ, $H_\rho \cup T_{H_\rho} \models \phi$ and, in particular, $\boldsymbol{H} \cup T_{\boldsymbol{H}} \not\models \phi$. Since $H_0 \subseteq \boldsymbol{H}$, we conclude that $H_0 \cup T_{\boldsymbol{H}} \not\models \phi$ and, hence, $\phi \notin \boldsymbol{T}$. □

Thus, $\boldsymbol{H}$ contains everything we can know about truth at all. So the fact that principles like 1-4 are not in $\boldsymbol{H}$ is a major problem for minimalism.

Can this situation be ameliorated? Yes, it can. It is well known that other fixed points of the supervaluational scheme, like the ones we have presented before —created using restrictions on the candidate extensions of the truth predicate— are much better behaved with respect to principles like 1-4. So the question is whether Horwich's construction can be manipulated so that the fixed point we get at the end be stronger than $\boldsymbol{H}$; ideally $\boldsymbol{VF}^{mc}$. This manipulation, though, should be made following independent reasons beyond the fact that $\boldsymbol{H}$ does not contain principles 1-4. This can be done, if we have in mind Horwich's position in front of the Liar paradox.

Let us begin by asking ourselves which model for the expanded language is the *actual* model; which model captures the actual world. First, since Horwich's position in front of the Liar defends, as we have seen in Section 6.2, that any sentence (in particular the Liar) is such that it is true or it is false, the extension of the truth predicate in the actual model will have to be, at least, complete; there will not be undecided cases of being true.

On the other hand, as we saw in section 6.1, Horwich adopts classical logic, which means that not only all sentences are either true or false, but also that it is not the case that they are true and false. Hence, it seems natural to expect from the extension of the truth predicate to be consistent; that is, no sentence is both true and false. All this means that the extension of the truth predicate in the actual model should be maximally consistent. On the other hand, by factivity of knowledge, it is reasonable to expect the extension of the truth predicate in the actual model to be a maximally consistent superset of $\boldsymbol{H}$.

These considerations naturally suggest to restrict our attention, in general, to models whose extension of the truth predicate is maximally consistent. Hence, we can bring this restriction to the consequence relation used in the construction.

Following this line of thought I will call a model $\mathcal{M}$ for $\mathcal{L}^+$ mc-acceptable, in symbols $\mathcal{M}_{mc}$ if, for every $\phi \in \mathcal{L}^+$, either $\ulcorner\phi\urcorner$ or $\ulcorner\neg\phi\urcorner$ belong to $Tr^{\mathcal{M}_{mc}}$ but not both. We can restrict, then, logical consequence to mc-acceptable models; that is, given a set of sentences Γ and a sentence α of $\mathcal{L}^+$, $\Gamma \models_{mc} \alpha$ if, and only if, for every mc-acceptable model $\mathcal{M}_{mc}$, if $|\Gamma|_{\mathcal{M}_{mc}} = 1$ then $|\alpha|_{\mathcal{M}_{mc}} = 1$.

The definition of the new series will be the same but substituting the unrestricted consequence relation by $\models_{mc}$. Lemma 6.4.1 can be proved in the same way (essentially it uses the fact that, if $\phi \in A$, for any sentence ϕ and set of sentences A, then $A \models_{mc} \phi$) and, hence, a fixed point of the construction, let us call it $\boldsymbol{H}^{mc}$, will exist. Lemmas 6.4.7 and 6.4.8 can easily

be proved for $\boldsymbol{H}^{mc}$ and $\boldsymbol{VF}^{mc}$ to show that $\boldsymbol{H}^{mc} = \boldsymbol{VF}^{mc}$.

Now all the instances of laws 1-4 are validated in $\boldsymbol{H}^{mc}$ and we have found some independent reasons —that is, Horwich's stance in front of the Liar— to motivate the adoption of this stronger fixed point.

Not everything are good news, though. Recall that some of the instances of the T-schema are not in $T_{\boldsymbol{H}}$, which means that the utility of the truth predicate, as presented in section 6.2, is seriously impaired. Moreover, this can be known from inside the model. For notice that, given $\lambda \leftrightarrow \neg Tr^{\ulcorner}\lambda^{\urcorner}$, then the $Tr^{\ulcorner}\lambda^{\urcorner}$-instance of the LEM, that is, $Tr^{\ulcorner}\lambda^{\urcorner} \vee \neg Tr^{\ulcorner}\lambda^{\urcorner}$, is equivalent to $\neg(Tr^{\ulcorner}\lambda^{\urcorner} \leftrightarrow \lambda)$. Now, since both $\lambda \leftrightarrow \neg Tr^{\ulcorner}\lambda^{\urcorner}$ and $Tr^{\ulcorner}\lambda^{\urcorner} \vee \neg Tr^{\ulcorner}\lambda^{\urcorner}$ are theorems, we have that $\neg(Tr^{\ulcorner}\lambda^{\urcorner} \leftrightarrow \lambda)$ is also a theorem and, hence, it is in $\boldsymbol{H}$ and in $\mathcal{T}$. Hence, it is not only conceptually possible to know that the Liar-instance of the T-schema is just false, but this is a fact about pure truth. This, of course, is hardly surprising; any theory of truth that solves the Liar paradox by keeping classical logic and restricting the T-schema will suffer from similar problems.[20]

6.5 Some Objections

It's time now to take stock. We have seen in section 6.3.2 that Horwich (1998b) proposes two conditions that the restriction of the T-schema should satisfy, the Maximality and the Specification constraints. We already saw that the former alone is not enough to specify a particular maximal consistent set of instances of the T-schema. We have seen, in the last section, Horwich's proposal with respect to the latter. Both conditions, though, are not yet met, even in the case where we take $\boldsymbol{H}^{mc}$ as the set of determinately true sentences. We saw the reason in chapter 5: consider the Truth Teller, the sentence τ such that $\tau \leftrightarrow Tr^{\ulcorner}\tau^{\urcorner}$. Notice that $\tau \notin \boldsymbol{H}^{mc}$ and, hence, although $\tau \leftrightarrow Tr^{\ulcorner}\tau^{\urcorner} \in \mathcal{T}$ —for the Diagonal Lemma shows that it is a theorem— and, hence, we can know that the τ-instance of the T-schema is true, we cannot know whether τ itself is true or not. Notice now that it would be perfectly safe to introduce τ already at H_0. In this case τ would obviously be at the fixed point resulting of the construction which would be a superset of $\boldsymbol{H}^{mc}$. As Kripke (1975) showed, there will be many maximal incompatible fixed points extending $\boldsymbol{VF}$ and we will not have any means to decide which one to adopt as the root of our truth theory.

In any case, suppose that we succeeded in singling out one unique fixed point containing sentences like τ. In this case, we would be able to know τ,

[20] See, for example, the discussion in Field (2008, p. 119).

because it would be in the fixed point, which, as we saw, contains everything we can know about truth. But mere desire for maximality seems a poor reason to achieve knowledge about τ or similar sentences; that is, the maximality constrain would give us knowledge about many sentences in a way that seems too way arbitrary. Having all these considerations into account, hence, it seems more reasonable to just abandon maximality and retain only the specification constrain.

We have also seen, in the last sections, the constructive specification Horwich proposes. Its result turned out to be identical to the minimal fixed point of Kripke's construction using the supervaluationist schema without restrictions on the candidate extensions of the truth predicate. Since the result of this construction was too weak, we presented a way to strengthen it in order to obtain the fixed point of Kripke's construction using the supervaluationist schema, but now with the condition of maximal consistency for the candidate extensions of the truth predicate. It is worth asking ourselves, then, why should we need Horwich's construction at all. Horwich (2010a) presents his construction as one that "squares with minimalism" (Horwich 2010a, p. 92, ft. 12) in the sense that it does not use compositional principles for truth, which are seen as being at odds with minimalism. This is so because Minimalism understands truth via the T-schema and not via compositional principles *à la* Tarski. Indeed, Horwich rejects Kripke's construction based on the strong Kleene scheme because it "invokes Tarski-style compositional principles" (Horwich 2010a, p. 92, ft. 12). The first thing we have to take into account is whether the supervaluationist version of the Kripke's construction is also invoking compositional principles. In this case it seems reasonable to expect that Horwich would also refuse to accept it, because even if the supervaluational scheme assigns the semantic value on the grounds of ways of making the truth predicate precise, in each of this ways, the semantic value of the sentences is achieved by compositional rules. This discussion clearly suggests, then, that Horwich takes the construction to be something more than a mere technical device to determine which instances of the T-schema are in the minimalist truth theory.

Horwich's construction, though, heavily relies on the notion of groundedness; he himself claims that "a good solution to the liar paradox should articulate 'grounding' constraints [...] on which particular instances of [the T-schema] are axioms" (Horwich 2010a, p. 91). But then, given that, as we concluded, the construction is not a mere technical device and has to satisfy strict deflationist constraints, using the notion of groundedness might seem to be at odds with Horwich's deflationist view about truth. After all, depending on how we understand the notion of groundedness, if it is constitutive of the notion of truth, it is no longer the case that "our commitment to [the

T-schema] accounts for everything we do with the truth predicate" (Horwich (1998b, p. 121)) and, hence, it is no longer the case that the T-schema implicitly defines it.

Horwich may have a way out of this situation; he might understand the construction of $\boldsymbol{VF}$ as a model of how truth claims are explained by some things in the world being in a certain way. Let me elaborate on that. Horwich admits that the correspondence intuition can be accommodated into Minimalism. As we saw in chapter 1, we can loosely characterize correspondence theories of truth as defending that being true consists in corresponding to facts. Although Horwich does not endorse, obviously, this characterization, he claims that "we might hope to accommodate much of what the correspondence theorist wishes to say without retreating an inch from our deflationary position" (Horwich 1998b, p. 104). The idea is that

> It is indeed undeniable that whenever a proposition or an utterance is true, it is true *because* something in the world is a certain way — something typically external to the proposition or the utterance. For example,[...]
>
> <Snow is white> is true *because* snow is white.

Thus, claims Horwich, the fact that snow is white is explanatorily prior to the fact that <Snow is white> is true, in the same way that the basic laws of nature and the initial state of the universe are explanatorily prior to the fact that snow is white.

I think that it is reasonable to understand the construction of $\boldsymbol{H}$ as an epistemic model of the relation of explanatory dependence between truth ascriptions and the extra-semantic facts. In this epistemic reading of groundedness, the grounded sentences, those in $\boldsymbol{H}$, are the ones that can be explained —and, hence, the ones that can be known— given how the world is and given the appropriate instances of the T-schema.

The problem is that, as we have seen, $\boldsymbol{H}$ is too weak. We were able to strengthen it by taking into account only the models in which the extension of the truth predicate was a maximally consistent set. Can we understand the construction of $\boldsymbol{H}^{mc}$ as an epistemic model of how we are able to explain truth ascriptions? I do not think so, at least not without begging the question. The problem is that we cannot use the claim that the extension of the truth predicate is a maximally consistent set in the construction unless it is something that we know, that is, something in $\boldsymbol{H}^{mc}$. But we cannot even construct $\boldsymbol{H}^{mc}$ without supposing that the extension of the truth predicate is maximally consistent. Thus, in building maximal consistency into the con-

struction we are going beyond what is given by the T-schema and, hence, we are abandoning deflationists views about truth.

6.6 Solving the Paradoxes

In this chapter we have seen how Horwich is able to give an uniform explanation of why the Liar and the *Sorites* are paradoxical. He uses an epistemic notion of indeterminacy according to which the use properties of the vague terms and 'true' make it impossible to know the semantic value of some of their ascriptions. Horwich's proposal is, hence, a common solution to the Liar and the *Sorites*, for it shows how the source of their paradoxicality are of the same nature.

It remains to see whether Horwich's epistemic approach is a strong common solution. I think the answer is no; I would say that, in the case of vagueness, the prevention is inexistent (or just vacuous), while in the case of truth a sophisticated construction is needed in order to accommodate the Liar within the intended theory of truth. Hence, Horwich is proposing a weak common solution to the Liar and the *Sorites* paradoxes.

CHAPTER

SEVEN

HARTRY FIELD: TOWARDS A CONDITIONAL FOR THE LIAR AND THE SORITES

Hartry Field has recently endorsed what can be seen as an effort to put forward a common solution to the Liar and the *Sorites*. In this chapter I will present how I understand to be the *paracomplete project*. Such a project consists in obtaining a paracomplete logic, stronger than *SK*, that can support a truth predicate and block the *Sorites*.

I will introduce some more reasons why, according to Field, we should reject LEM to cope with vagueness and truth and we will see what sense of rejection is used when stating that some instances of LEM should be rejected.

Finally, I will put forward a suggestion of where I think the investigation could go by giving a model, based on Field's work, for a first-order language with two conditionals, a truth predicate and vague predicates.

7.1 The Liar, the *Sorites* and Indeterminacy

Hartry Field has defended in a number of places (see Field 2003b, Field 2003c and Field 2008) that the Liar and the *Sorites* paradoxes are somehow connected and that, consequently, should be treated in a unified way.

Truth and vagueness are connected, according to Field, in the sense that both give rise to questions for which there is no fact of the matter about their answers. Such lack of factivity would be prompted, claims Field, by

> the fact that the standard means for explaining 'true' (namely, the [T-schema]) fails to uniquely determine the application of the term to certain sentences (the "ungrounded" ones); and this seems to be just

> the sort of thing that gives rise to other sorts of vagueness. (Field 2003c, p. 308)

Compare the following claims (and suppose I am a borderline case of being tall):

1. Sergi Oms is tall.
2. The Liar is true.
3. Tarski's mother weighed less than 70 kilos when she died.

As Field (2003b, p. 461) points out, we have different attitudes towards claims like 1 and 2, on the one hand, and 3 on the other. Field proposes to consider 3 as a "perfectly factual claim" while 1 and 2 as non factual claims. One first obvious objection to 1 being non factual is the existence of the *Sorites* paradox, which seems to conclude, precisely, that there is a fact of the matter whether Sergi Oms is tall. To see this, consider, again, the line-drawing *Sorites* introduced at page 16:

Line-drawing *Sorites*

Pa_1

$\neg Pa_n \ (n > 1)$

$\forall x(Pa_x \vee \neg Pa_x)$

Hence, there is a z $(1 \leq z < n)$ such that Pa_z and $\neg Pa_{z+1}$

If we interpret P as 'tall' and $a_1, \ldots, a_x$ as the appropriate series, the resulting argument seems to show that there is a fact of the matter as to which height exactly divides tall people from non tall people and, hence, there is a fact of the matter as to whether Sergi Oms is tall or not. Field's response to that is to reject LEM and, hence, block the above argument at the third premise. As a matter of fact, Field proposes to capture the difference between 1-2 and 3 with the idea that only in the case of 3 would we accept its LEM-instance.

As far as I can see, Field's strategy can be understood of as noting, first, that vague ascriptions to borderline cases are non factual, which amounts to the rejection of its instances of LEM. Next, since what prompts such non factuality seems to be the fact that our linguistic practices fail to determine a unique extension for vague predicates and, moreover, this seems to be the same that happens with truth in the case of truth paradoxes, it is sensible

to adopt non factuality for certain truth claims too. Such non factuality will also involve the rejection of certain instances of LEM. This, together with the mentioned fact, at the beginning of chapter 5, that restricting LEM blocks the Liar argument makes the case for adopting a paracomplete logic to cope with both the Liar and the *Sorites*.

7.1 Rejecting LEM

If that is so, Field needs to clarify, first, what it is to *reject* certain instances of LEM. As he notices (see, for instance, Field 2003c, p. 275), *reject* cannot mean *deny*, in the sense of *accepting the negation*. If this were so, then rejecting the LEM-instance of 1 would be the same as accepting:

4. ¬((Sergi Oms is tall) ∨¬ (Sergi Oms is tall))

But if we deny a disjunction it is because we are disposed to deny each one of the disjuncts, which means that if we accept 4, we are committed to assert both of the following (the denials of both disjuncts of 4):

5. ¬(Sergi Oms is tall)

6. ¬¬(Sergi Oms is tall)

which is a contradiction.[1] Nor will it help to interpret *reject* as *not true*, for, since we are supposing that any claim is equivalent to its truth ascription, asserting that the LEM-instance of 1 is not true is the same as asserting 4 and, hence, we are driven again to a contradiction.

Rejection cannot be *nonacceptance* either. To see this consider again claims 1 and 3 above; I am willing to reject 1 and its negation —due to its non-factuality— but, although I do not accept neither 3 nor its negation, I do not reject them, for, given the factuality of 3 and its negation, rejecting one of them would commit me to accepting the other. This shows that I do not reject 3, in spite of the fact that I do not accept it. Consequently, *rejection* is stronger than mere *nonacceptance*.

This discussion suggests that explaining *rejection* in terms of other notions is not a sensible strategy. Field's way out of this situation consists in considering *rejection* as the dual of *acceptance* and, hence, as a primitive notion. In order to illuminate our understanding of both *rejection* and *acceptance*, Field offers a model of how they relate to each other by idealizing epistemic attitudes and supposing that rational agents have numerical

[1] In this chapter I am setting aside the possibility of embracing a paraconsistent view.

degrees of belief represented by a function P ranging over the real interval $[0,1]$. The function P is typically taken to obey the laws of classical probability. In relation to that, we need to notice that if

(i) rejecting a claim ϕ implies that $P(\phi) < 1$,

(ii) we want to reject claims of the form $\phi \vee \neg\phi$, and

(iii) $P(\phi \vee \neg\phi) = P(\phi) + P(\neg\phi)$ (as classical probabillty entails),

then we need to weaken the classical probability law,

(CL) $P(\phi) + P(\neg\phi) = 1$,

to

(CL_w) $P(\phi) + P(\neg\phi) \leq 1$

That means, unsurprisingly, that classical probability must be revised in order to accommodate rejection of LEM. How, then, the weakening of (CL) to (CL_w) helps enlighten the relation between *acceptance* and *rejection*? In the present framework, *acceptance* of ϕ can be naturally understood of as having a high enough degree of belief in ϕ; in other words, as having a degree of belief in ϕ over a certain threshold $\tau > 1/2$, that is, $P(\phi) \geq \tau$. We can then see *rejection* as the dual of *acceptance* so that rejecting ϕ is having a degree of belief in ϕ lower that the co-threshold $1 - \tau$. It can be seen, now, that if (CL) is accepted, then rejecting ϕ just amounts to accepting $\neg\phi$, but if it is weakened to (CL_w), then we can reject something without accepting its negation, which means that *rejection* is weaker than *denial*. There can be now a ϕ such that we reject both ϕ and $\neg\phi$. Also, as desired, not accepting ϕ does not imply rejecting it, so that *rejection* is stronger than *nonacceptance*. This makes *rejection* a notion in between *denial* and *nonacceptance*.

We can now express the difference between factual claims like 3 and non factual claims like 1 and 2 by using our degree of belief in their LEM-instances (supposing, for simplicity, that 1 and 2 are clear cases of indeterminate claims):

$$P((\text{Sergi Oms is tall}) \vee \neg(\text{Sergi Oms is tall})) = 0$$

$$P((\text{The Liar is true}) \vee \neg(\text{The Liar is true})) = 0$$

$$P((\text{Tarski's mother weighed less than 70 kilos when she died}) \vee \neg(\text{Tarski's mother weighed less than 70 kilos when she died})) = 1$$

What characterizes an indeterminate sentence ϕ, then, is that we reject both ϕ and $\neg\phi$, which amounts to rejecting $\phi \vee \neg\phi$.

7.1 Paracomplete Logics

The discussion in the previous section together with the already noticed fact at the beginning of the chapter 5 that rejecting LEM allows us to block the Liar paradox gives us enough reasons to try to solve both the *Sorites* and the Liar with the use of a paracomplete logic. One natural candidate for such a logic would be SK as introduced in chapter 5. We already saw, though, that SK suffers from two grave problems; first, the logic (specially the logic governing the conditional) is too weak as it does not even validate principles like $\phi \rightarrow \phi$ and, second, it suffers from revenge problems, which can be seen as an inability to express a suitable operator of determinacy.

Part of the paracomplete project endorsed by Field, then, is to obtain a stronger paracomplete logic that can satisfy the Intersubstitutivity Principle of Truth (IP), introduced at page 6, which, given that the logic is supposed to be strong enough to validate $\phi \rightarrow \phi$, amounts to satisfying the T-schema. Since the main culprit of the logical weakness of SK seems to be the material conditional, we can try to add a suitable conditional to that logic, while preserving the good behavior of the truth predicate. Moreover, having in mind the discussion of the previous section, such logic should be able to block the *Sorites* paradox as well.

There is still another caveat we need to have in mind with respect to the conditional we are seeking. Recall Curry's paradox as introduced at chapter 1:

1. $\gamma \leftrightarrow (Tr\ulcorner\gamma\urcorner \rightarrow \phi)$	(Diagonal Lemma)
2. $Tr\ulcorner\gamma\urcorner \rightarrow (Tr\ulcorner\gamma\urcorner \rightarrow \phi)$	(γ-instance of the T-schema and logic on 1)
3. $Tr\ulcorner\gamma\urcorner \rightarrow \phi$	(contraction —$\psi \rightarrow (\psi \rightarrow \chi) \vdash \psi \rightarrow \chi$— on 2)
4. $(Tr\ulcorner\gamma\urcorner \rightarrow \phi) \rightarrow Tr\ulcorner\gamma\urcorner$	(γ-instance of the T-schema and logic on 1)
5. $Tr\ulcorner\gamma\urcorner$	(*Modus Ponens* on 3 and 4)
6. ϕ	(*Modus Ponens* on 3 and 5)

Given this paradox, if we want to keep the T-schema, the most obvious culprit we are left with is the contraction rule on step 3. Consequently, we are looking for a conditional, strong enough to validate $\phi \rightarrow \phi$ (and other laws we would expect of a conditional), weak enough to fail to validate contraction and such that can be added to SK in a way that preserves (IP). One natural candidate is the conditional of Łukasiewicz 3-valued logic $\rightarrow_3$:

$\rightarrow_3$	0	1	1/2
0	1	1	1
1	0	1	1/2
1/2	1/2	1	1

If we add $\rightarrow_3$ to *SK* we obtain Łukasiewicz 3-valued logic, Ł$_3$. It is easy to see that contraction is not valid in Ł$_3$. This move, though, will not do for, with $\rightarrow_3$, we can define the biconditional $\phi \leftrightarrow_3 \psi$ as $(\phi \rightarrow_3 \psi) \wedge (\psi \rightarrow_3 \phi)$ which is, precisely, one of the biconditionals that, at page 78, we saw cannot be added to *SK* together with a truth predicate satisfying (IP).

Still, following Field (2008), we can see that the way how Ł$_3$ copes with Curry's paradox might be enlightening. Recall that a Curry sentence is a sentence γ that is equivalent to $Tr^{\ulcorner}\gamma^{\urcorner} \rightarrow \phi$. Suppose we are in the worst of the cases and ϕ is a contradiction. Notice, first, that in Ł$_3$ we can assign the same value in a consistent way to γ and $Tr^{\ulcorner}\gamma^{\urcorner}$. For clarity let us use $|\phi|^3$ as a symbol for the semantic value of ϕ under Ł$_3$. The Ł$_3$ conditional can be defined now in a more compact way:

$$|\phi \rightarrow_3 \psi|^3 = \begin{cases} 1 \text{ if } |\phi|^3 \leq |\psi|^3 \\ 1 - (|\phi|^3 - |\psi|^3) \text{ if } |\phi|^3 > |\psi|^3 \end{cases}$$

Ł$_3$ conditional, thus, has semantic value 1 when the value of the antecedent is less than or equal to the value of the consequent; that is, when, so to speak, there is no "loss of truth". When the value of the antecedent is strictly greater than the value of the consequent, so that there is "loss of truth", the value of the conditional is 1 minus the "lost truth".

Now we can see how we can consistently assign the same value to γ and $Tr^{\ulcorner}\gamma^{\urcorner}$. Notice that, since $|\phi|^3 = 0$, then $|\gamma|^3 = |Tr^{\ulcorner}\gamma^{\urcorner} \rightarrow \phi|^3 = 1 - |Tr^{\ulcorner}\gamma^{\urcorner}|^3$. Then, what we need is $|\gamma|^3 = |Tr^{\ulcorner}\gamma^{\urcorner}|^3 = 1 - |Tr^{\ulcorner}\gamma^{\urcorner}|^3$, which can be accomplished by assigning ½ to γ. Goal achieved. Unfortunately, Curry's problems do not end here. Consider the sentence γ_1 equivalent to $Tr^{\ulcorner}\gamma_1{}^{\urcorner} \rightarrow \gamma$, where γ is the previous Curry's sentence. Again, we want to consistently assign the same value to γ_1 and $Tr^{\ulcorner}\gamma_1{}^{\urcorner}$. Suppose it can be done, so that $|\gamma_1|^3 = |Tr^{\ulcorner}\gamma_1{}^{\urcorner}|^3$. Then, $|\gamma_1|^3 = |Tr^{\ulcorner}\gamma_1{}^{\urcorner} \rightarrow \gamma|^3 = |\gamma_1 \rightarrow \gamma|$, which, given the rules of the conditional implies that $|\gamma_1|^3 = 3/4$. Since this is an absurdity —we do not have such semantic value in our logic—, we conclude that our previous supposition was false and that, consequently, we cannot consistently assign the same value to γ_1 and $Tr^{\ulcorner}\gamma_1{}^{\urcorner}$. Still, we could try adding the semantic value 3/4 to tour logic (and its dual 1/4) so that $|\gamma_1|^3 = |Tr^{\ulcorner}\gamma_1{}^{\urcorner}|^3 = 3/4$. But a predictable difficulty appears when we consider the sentence γ_2 equivalent to $Tr^{\ulcorner}\gamma_2{}^{\urcorner} \rightarrow \gamma_1$, which forces us to add the semantic value 7/8 (and its dual 1/8) to the logic. This process can go on indefinitely, so that the requirement that $|\gamma_n|^3 = |Tr^{\ulcorner}\gamma_n{}^{\urcorner}|^3$ is never met, at least for a finite semantic value space. Now an idea comes naturally to mind:

why do not consider an infinite valued semantics? And the most natural candidate is the Łukasiewicz continuum valued logic $Ł_\infty$.

As a matter of fact, $Ł_\infty$ is a very promising candidate for the paracomplete logic we are seeking. First, a truth predicate satisfying (IP) can be added to Łukasiewicz continuum valued sentential logic.[2] Second, $Ł_\infty$ is one of the paradigmatic examples of a many-valued approach to vagueness. I will not pursue here, though, the compellingness of $Ł_\infty$ applied to vagueness, for it will be enough, for my purposes, to show that, in fact, $Ł_\infty$ cannot support a truth predicate.[3]

Let us introduce, first, $Ł_\infty$, which is a generalization of $Ł_3$ that uses as semantic values, as I said, the real interval $[0,1]$ (for our present purposes we can think of 0 as "false" and 1 as "true"). Given a first-order language $\mathcal{L}$, an $Ł_\infty$ model $\mathcal{M}$ will consist of a domain D together with an interpretation for each predicate (a function from the appropriate cartesian product of D to the real interval $[0,1]$) in the language so that the values of atomic sentences are settled. I will use $|\phi|^\infty_{\mathcal{M}}$ as a symbol for the semantic value of ϕ in $Ł_\infty$ under the model $\mathcal{M}$. Then, given a model $\mathcal{M}$, for any sentences ϕ and ψ, the values for complex sentences are determined as follows:

1. $|\neg\phi|^\infty_{\mathcal{M}} = 1 - |\phi|^\infty_{\mathcal{M}}$
2. $|\phi \vee \psi|^\infty_{\mathcal{M}} = \max\{|\phi|^\infty_{\mathcal{M}}, |\psi|^\infty_{\mathcal{M}}\}$
3. $|\phi \wedge \psi|^\infty_{\mathcal{M}} = \min\{|\phi|^\infty_{\mathcal{M}}, |\psi|^\infty_{\mathcal{M}}\}$
4. $$|\phi \to \psi|^\infty_{\mathcal{M}} = \begin{cases} 1 & \text{if } |\phi|^\infty_{\mathcal{M}} \leq |\psi|^\infty_{\mathcal{M}} \\ 1 - (|\phi|^\infty_{\mathcal{M}} - |\psi|^\infty_{\mathcal{M}}) & \text{if } |\phi|^\infty_{\mathcal{M}} > |\psi|^\infty_{\mathcal{M}} \end{cases}$$
5. $|\exists x\phi|^\infty_{\mathcal{M}} = \sup\{|\phi(d/x)|^\infty_{\mathcal{M}} : d \in D\}$
6. $|\forall x\phi|^\infty_{\mathcal{M}} = \inf\{|\phi(d/x)|^\infty_{\mathcal{M}} : d \in D\}$

In the last two clauses $\phi(d/x)$ is the result of replacing all free occurrences of x in ϕ with d.[4]

[2] See the Appendix of chapter 4 in Field (2008) for a proof.

[3] For a defense of $Ł_\infty$ applied to vagueness see Machina (1976) or Smith (2008) and for some criticisms see Keefe (2000, chapter 4).

[4] As I did before, I am supposing that every member d of the domain serves as a name of itself.

As Restall (1994) shows, the T-schema cannot hold in Ł$_\infty$. In order to see this, let us first introduce a new logical connective, $\circ$, called *fusion* (or, sometimes, also *strong conjunction* or *T-norm conjunction*):

7. $\phi \circ \psi$ is defined as $\neg(\phi \rightarrow \neg\psi)$

I will use ϕ^n to denote the n-fold fusion of ϕ with itself; thus, $\phi^3 = \phi \circ (\phi \circ \phi)$, $\phi^4 = \phi \circ (\phi \circ (\phi \circ \phi))$, and so on. Now we can informally sketch the main argument in Restall (1994). I will use the fact that if a formula ϕ is such that $|\phi|^\infty_{\mathcal{M}} < 1$ in a structure $\mathcal{M}$, then for some m, $|\phi^m|^\infty_{\mathcal{M}} = 0$ (see Restall 1994, p. 2 for the proof). Consider next the following series of sentences:

$A_0 = \neg\forall x > 0\ A_x$ is true

$A_{n+1} = A_0^{n+1}$

Now we can reason informally as follows. Suppose A_0 is true. Then, by construction of each A_{n+1} (fusion iterations), all of them are true, which means that A_0 is not true after all. Since we have reached a contradiction, we conclude that A_0 is not true. Then, though, by the previous fact about fusion iterations, there will be an m such that A_m will be false, which means that A_0 is true. Contradiction.[5] This shows that, unfortunately, Ł$_\infty$ is not a suitable logic for truth.

We have seen, now, two logics that have failed as the paracomplete logic we were looking for. *SK* was too weak, whereas Ł$_\infty$ seems to be too strong, in the sense that, as we have seen, it leads to ω-inconsistency when combined with a truth predicate.

Hartry Field has tried to devise some logics for truth stronger than *SK* and weaker than Ł$_\infty$ (see Field 2003a,c, 2007, 2008 and, more recently, Field 2014, 2016). As Restall (1994) points out, the logic called *CK* (or, sometimes,

[5]More precisely, Restall (1994) shows that adding the T-schema to Ł$_\infty$ yields ω-inconsistency.

Actually, this can be seen as a revenge problem. Fusion can be used to defined a definitely operator $D\phi = \phi \circ \phi$ to express the indeterminacy of the Liar, so that $\neg D\lambda \wedge \neg D\neg\lambda$ comes out true. We can then define a new liar sentence, λ_1, equivalent to $\neg DTr^\ulcorner\lambda_1\urcorner$ and, again, express its indeterminacy with $\neg DD\lambda_1 \wedge \neg DD\neg\lambda_1$. This process can go on for sentences λ_n each one equivalent to $\neg D^n Tr^\ulcorner\lambda_n\urcorner$ whose indeterminacy can be expressed with the use of a D^n operator. What Restall (1994) shows is that the process collapses at D^ω. See Hajek, Paris, and Sheperdson (2000) for a generalization of this result.

also *RWK*) is a good place to start; it is a paracomplete logic, stronger than *SK* and slightly weaker than Ł$_\infty$. *CK* can be axiomatized, following Priest (2008), as follows:

1. $\phi \rightarrow \phi$
2. $\phi \rightarrow (\phi \vee \psi)$
3. $(\phi \wedge \psi) \rightarrow \phi$
4. $(\phi \wedge (\psi \vee \chi)) \rightarrow ((\phi \wedge \psi) \vee (\phi \wedge \chi))$
5. $((\phi \rightarrow \psi) \wedge (\phi \rightarrow \chi)) \rightarrow (\phi \rightarrow (\psi \wedge \chi))$
6. $((\phi \rightarrow \chi) \wedge (\psi \rightarrow \chi)) \rightarrow ((\phi \vee \psi) \rightarrow \chi)$
7. $\neg\neg\phi \rightarrow \phi$
8. $(\phi \rightarrow \neg\psi) \rightarrow (\psi \rightarrow \neg\phi)$
9. $(\phi \rightarrow \psi) \rightarrow ((\psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi))$
10. $(\phi \rightarrow \psi) \rightarrow ((\chi \rightarrow \phi) \rightarrow (\chi \rightarrow \psi))$
11. $\phi \rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$
12. $\phi \rightarrow (\psi \rightarrow \phi)$
13. $\forall x\phi \rightarrow \phi(x/t)$
14. $\forall x(\phi \rightarrow \psi) \rightarrow (\phi \rightarrow \forall x\psi)$
15. $\forall x(\psi \rightarrow \phi) \rightarrow (\exists x\psi \rightarrow \phi)$
16. $\forall x(\phi \vee \psi) \rightarrow (\phi \vee \forall x\psi)$[6]

R1 $A, A \rightarrow B \models B$

R2 $A, B \models A \wedge B$

R3 If $\models A(x)$, then $\models \forall xA$

[6]In 14, 15 and 16 x is not free in A.

Field has offered a family of logics weaker than *RWK* to which a truth predicate satisfying (IP) can be added. He has showed their consistency by offering models for a language with a truth predicate. Unfortunately, as we are going to see, the logics Field has presented are still too weak as they do not satisfy some principles that we would like to be satisfied. In this chapter I want to explore the possibility of having two conditionals in the language in order to have a more satisfactory logic. I will use a conditional intended to capture certain analytic relations that occur between sentences of a language with a truth predicate and vague predicates. In particular, I want to see whether this conditional can capture the relation of entailment and the penumbral connections.

7.2 The Construction

7.2 The First Conditional

I will proceed with the construction that defines the first conditional. Suppose we have a language $\mathcal{L}$ without conditionals, suitable to express canonical names for its own sentences and we want to extend it to a new language, $\mathcal{L}^+$, with a truth predicate, Tr and two conditionals, $\rightarrow$ and $\Rightarrow$. Let us define models $\mathcal{M} =< W^{\mathcal{M}}, D^{\mathcal{M}}, I^{\mathcal{M}} >$ for $\mathcal{L}$ as ordered triples with a set of points, a domain and an interpretation function. $W^{\mathcal{M}}$ is a set of three valued points (the semantic values will be 0, ½ and 1) and a $I^{\mathcal{M}}$ is a function that gives the denotations of the non-logical terminology at each point. We have a single non-empty set, $D^{\mathcal{M}}$, as the domain of discourse and $I^{\mathcal{M}}$ gives us the appropriate interpretation for constants and function expressions in the usual way. I am specially interested here in the case of the predicates, hence I will skip the details concerning the other non-logical expressions. Furthermore I will simplify and consider only unary predicates; thus, for each predicate the function $I^{\mathcal{M}}$ yields, at each point, the extension and the anti-extension of the predicate, two disjoint and not necessarily exhaustive subsets of $D^{\mathcal{M}}$. As usual, the sentences assigning a given predicate to an object of the extension will receive the semantic value 1, the sentences assigning a given predicate to an object of the anti-extension will receive the semantic value 0, and the sentences assigning a given predicate to an object which belongs neither to the extension nor to the anti-extension will receive the semantic value ½.

Let us denote, for a predicate P of $\mathcal{L}$ and a point w, the extension of P at w as P^+_w and the anti-extension of P at w as P^-_w. The semantic value of a sentence A at a point w will be denoted by $\mid A \mid_w$ together with other subscripts that I will introduce shortly. I will not mention the model unless

it is necessary.

Next, we can define an order for the points of W: for all w_1 and $w_2 \in W$, $w_1 \leq w_2$ iff for each predicate P, $P^+_{w_1} \subseteq P^+_{w_2}$ and $P^-_{w_1} \subseteq P^-_{w_2}$.

Let us define how the logical constants assign truth values to sentences of $\mathcal{L}^+$ and how the truth predicate and the conditionals are to be understood. The values of the conditionals are given by two functions, j and v. The function j for $\Rightarrow$ assigns to each $w \in W$ and each sentence of the form $\phi \Rightarrow \psi$ a value in $\{0, 1/2, 1\}$. Similarly, v is a function that assigns to each $w \in W$ and each sentence of the form $\phi \rightarrow \psi$ a value in $\{0, 1/2, 1\}$. I will write $j_u(\phi \rightarrow \psi)$ instead of $j(u, \phi \rightarrow \psi)$, and the same for v. The value of a sentence ϕ will depend on the following: a point $u \in W$, functions j and v for the value of the conditionals and a set of sentences X of $\mathcal{L}^+$ for the truth predicate (the extension of Tr):[7]

$$|Tr(\ulcorner\phi\urcorner)|_{\langle u,j,v,X\rangle} = \begin{cases} 1 & \text{iff } \phi \in X \\ 0 & \text{iff } \neg\phi \in X \\ 1/2 & \text{otherwise} \end{cases}$$

$$|\neg\phi|_{\langle u,j,v,X\rangle} = 1 - |\phi|_{\langle u,j,v,X\rangle}$$

$$|\phi \wedge \psi|_{\langle u,j,v,X\rangle} = \min\{|\phi|_{\langle u,j,v,X\rangle}, |\psi|_{\langle u,j,v,X\rangle}\}$$

$$|\phi \vee \psi|_{\langle u,j,v,X\rangle} = \max\{|\phi|_{\langle u,j,v,X\rangle}, |\psi|_{\langle u,j,v,X\rangle}\}$$

$$|\forall x\phi|_{\langle u,j,v,X\rangle} = \min\{|\phi(d/x)|_{\langle u,j,v,X\rangle} : d \in D\}$$

$$|\exists x\phi|_{\langle u,j,v,X\rangle} = \max\{|\phi(d/x)|_{\langle u,j,v,X\rangle} : d \in D\}$$

$$|\phi \Rightarrow \psi|_{\langle u,j,v,X\rangle} = j_u(\phi \Rightarrow \psi)$$

$$|\phi \rightarrow \psi|_{\langle u,j,v,X\rangle} = v_u(\phi \rightarrow \psi)$$

Given the semantic rules above and a model $\mathcal{M}$ for $\mathcal{L}$ we can construct a fixed point in the way we have already seen. First, we prove the following monotonicity principle.

Lemma 7.2.1 (Kripke 1975) *For all $\mathcal{M}$, j and v, if $X \subseteq X'$ then for any $u \in W$ and any sentence $\phi \in \mathcal{L}^+$, if $|\phi|_{\langle u,j,v,X\rangle}$ is an integral value (0 or 1), then $|\phi|_{\langle u,j,v,X\rangle} = |\phi|_{\langle u,j,v,X'\rangle}$.*

[7] For the quantifiers I will assume that all objects in the domain of discourse have as name the object itself. I will also use $\phi(d/x)$ as the result of replacing all free occurrences of x in ϕ with the term d.

Proof Suppose $X \subseteq X'$. The proof proceeds by induction on the complexity of ϕ. If ϕ is an atomic sentence of $\mathcal{L}$ the result is clear, for ϕ's semantic value is given by the interpretation function I.

Suppose, now, that $\phi = Tr^{\ulcorner}\psi^{\urcorner}$ and $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,v,X\rangle} = 1$. Then, by the semantic rules for Tr, $\psi \in X$ and, since $X \subseteq X'$ by supposition, $\psi \in X'$ and, consequently, $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,v,X'\rangle} = 1$. Suppose, next, $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,v,X\rangle} = 0$. Then, $\neg\psi \in X \subseteq X'$ and, hence, $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,v,X'\rangle} = 0$.

When $\phi = \neg\psi$ and $|\neg\psi|_{\langle u,j,v,X\rangle} = 1$, then $|\psi|_{\langle u,j,v,X\rangle} = 0$ and, by induction hypothesis, $|\psi|_{\langle u,j,v,X'\rangle} = 0$. Consequently, $|\neg\psi|_{\langle u,j,v,X'\rangle} = 1$. If $|\neg\psi|_{\langle u,j,v,X\rangle} = 0$ the result follows similarly. The other connectives are similar. Notice that the conditionals $\rightarrow$ and $\Rightarrow$ are treated as atomic sentences and, thus, the result follows for them trivially.

As we already have seen, the previous result implies the existence of a fixed point for the truth predicate.

Proposition 7.2.2 (Kripke 1975) *For any $\mathcal{M}$, j and v, there are X such that for every $u \in W$ and every sentence $\phi \in \mathcal{L}^{+}$, $|\phi|_{\langle u,j,v,X\rangle} = |Tr^{\ulcorner}\phi^{\urcorner}|_{\langle u,j,v,X\rangle}$.*

The proof uses a construction like the one in chapter 5 and the considerations of Theorem 6.4.2. In particular, for any $\mathcal{M}$, j and v there is a minimal fixed point, $\boldsymbol{K}$. Notice that the conditionals are still completely opaque to us, so that we cannot guarantee that the truth predicate will satisfy intersubstitutivity. But we can easily see a sufficient condition that the functions that govern the conditionals have to meet in order to guarantee the Intersubstitutivity for truth. Following Field (2016) let us introduce the following definition:

> A valuation j is *transparent* if, and only if, for any sentences ϕ and ψ, if ϕ^{*} is the result of substituting one or more occurrences of ψ in ϕ by $Tr^{\ulcorner}\psi^{\urcorner}$, then, for any $u \in W$, $j_u(\phi) = j_u(\phi^{*})$ (and the same for v).

Next we can see that, if j and v are transparent, then truth obeys intersubstitutivity.

Proposition 7.2.3 *For any sentences ϕ, ϕ^{*} and ψ such that ϕ^{*} is the result of substituting one or more occurrences of ψ in ϕ by $Tr^{\ulcorner}\psi^{\urcorner}$, any transparent j and v, and any $u \in W$, $|\phi|_{\langle u,j,v,\boldsymbol{K}\rangle} = |\phi^{*}|_{\langle u,j,v,\boldsymbol{K}\rangle}$.*

Proof The proof proceeds by induction on the depth of the embedding of the substituted occurrence of ψ in ϕ. First, if $\phi = \psi$, then $\phi^{*} = Tr^{\ulcorner}\psi^{\urcorner}$ and the result follows from the fact that $\boldsymbol{K}$ is a fixed point.

Suppose that $\phi = \theta_1 \vee \theta_2$. Then $|\phi|_{\langle u,j,v,\boldsymbol{K}\rangle} = \max\{|\theta_1|_{\langle u,j,v,\boldsymbol{K}\rangle}, |\theta_2|_{\langle u,j,v,\boldsymbol{K}\rangle}\}$. By induction hypothesis, $\max\{|\theta_1|_{\langle u,j,v,\boldsymbol{K}\rangle}, |\theta_2|_{\langle u,j,v,\boldsymbol{K}\rangle}\} = \max\{|\theta_1^*|_{\langle u,j,v,\boldsymbol{K}\rangle}, |\theta_2^*|_{\langle u,j,v,\boldsymbol{K}\rangle}\} = |\theta_1^* \vee \theta_2^*|_{\langle u,j,v,\boldsymbol{K}\rangle} = |\phi^*|_{\langle u,j,v,\boldsymbol{K}\rangle}$. The other logical constants different from the conditonals are proved similarly.

Finally, suppose that $\phi = \theta_1 \rightarrow \theta_2$. Then the transparency of v guarantees that $v_u(\phi) = v_u(\phi^*)$ so that $|\phi|_{\langle u,j,v,\boldsymbol{K}\rangle} = |\phi^*|_{\langle u,j,v,\boldsymbol{K}\rangle}$ (and the same for $\Rightarrow$). □

I will write $|\phi|_{\langle u,j,v\rangle}$ instead of $|\phi|_{\langle u,j,v,\boldsymbol{K}\rangle}$; it is clear now that, given a base model and functions j and v, we will find a minimal fixed point relative to them.[8]

We need now to construct appropriate transparent j and v to cope with the conditionals. In order to do that, Field (2003a, 2008, 2016) has used revision constructions based on the work in Gupta and Belnap (1993). Let us see how we can achieve the desired valuations for the conditionals.

First, consider a fixed given function j for all $\Rightarrow$-conditionals. We want to define a function v that will yield the functions v_u for all $\rightarrow$-conditionals and $u \in W$. Valuations v will depend now on the fixed j, so that this dependency might be made explicit with a subscript, nevertheless, for readability, I will drop such a subscript when the dependency is clear from the context. In order to achieve the desired function v we will construct a revision sequence in the sense of Gupta and Belnap (1993) and we will use its properties to obtain a transparent privileged v. The revision sequence will consist of a series of valuations defined over the class of all ordinals. The process starts with an initial valuation v_0 which assigns the value ½ to all the $\rightarrow$-conditionals. Then, we need to specify how to obtain new members of the process form the previous ones. I will denote revision sequences with (v_κ) and the stages of the sequences with v_κ, for some ordinal κ.Given a valuation v_κ for $\rightarrow$-Conditionals, I will write $|\phi|_{\langle u,j,\kappa\rangle}$ instead of $|\phi|_{\langle u,j,v_\kappa\rangle}$.

Now, we need to know how to construct a new valuation v_κ given the previous valuations. We will characterize each $v_{u,j,\kappa}$, for every $u \in W$. This will be done in the following way, where $u \in W$ and α is an ordinal:

$$v_{u,j,\alpha}(\phi \rightarrow \psi) = \begin{cases} 1 \text{ iff } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha)), & |\phi|_{\langle u,j,\gamma'\rangle} \leq |\psi|_{\langle u,j,\gamma'\rangle} \\ 0 \text{ iff } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha)), & |\phi|_{\langle u,j,\gamma'\rangle} = 1 \text{ and } |\psi|_{\langle u,j,\gamma'\rangle} = 0 \\ \text{½ otherwise.} \end{cases}$$

[8] So, properly speaking, $\boldsymbol{K}$ should have some kind of indexes pointing to this dependency, but, for readability, I will omit them.

It is easy to see that the sequence (v_κ) just defined preserves transparency and, hence, given that v_0 is transparent, all the members of (v_κ) are.

As I said, the sequence (v_κ) is a revision sequence as used in Gupta and Belnap (1993, pp. 167–168). According to Gupta and Belnap (1993), a rule of revision is an operation on a space of functions. In our case, each v_κ is a function from $\rightarrow$-conditionals and points in W to $\{0, 1/2, 1\}$ and the rule above tells us how to obtain each v_κ from the previous ones. In more detail, when κ is a successor ordinal $\sigma + 1$, v_κ depends on v_σ and, when κ is a limit ordinal, the value that v_κ assigns to a given $\rightarrow$-conditional will depend on whether the appropriate conditions stabilize in the series up to κ.

Revision sequences have very interesting properties. One of them is that there are valuations v_κ that appear arbitrarily late, that is, there are valuations v_κ such that for any ordinal σ, there is an ordinal θ, $\theta \geq \sigma$, such that $v_\kappa = v_\theta$. Let us call *cofinal* the ordinals κ such that v_κ has this property and $\text{COFIN}_{(v_\kappa)}$ the class of cofinal ordinals for valuations in the sequence (v_κ). As Gupta and Belnap (1993, p. 170) show, there is a least cofinal ordinal α, called *the initial ordinal* for (v_κ), such that for all ordinals σ, $\sigma \geq \alpha$, σ is cofinal.

Not all cofinal ordinals assign the same valuation to $\rightarrow$-conditionals; if they did, we would have a fixed point. So we need a privileged v_κ that we can use to define validity. In order to obtain the appropriate ordinal κ, we use the following theorem, adapted from the *Reflection Theorem* in Gupta and Belnap (1993).

Theorem 7.2.4 (Reflection theorem, Gupta and Belnap 1993) *There are limit ordinals $\Delta_{(v_\kappa)}$ (called* reflection ordinals for the sequence (v_κ)*) such that,*

(i) $\Delta_{(v_\kappa)} \in \mathit{COFIN}_{(v_\kappa)}$,

(ii) For any sentences ϕ and ψ in $\mathcal{L}^+$, any world $u \in W$ and any $d \in \{0, 1/2, 1\}$,

$[\forall \sigma \in \mathit{COFIN}_{(v_\kappa)},\ v_{u,\sigma}(\phi \rightarrow \psi) = d]$ *if, and only if,* $[(\exists \beta < \Delta_{(v_\kappa)})(\forall \gamma \in [\beta, \Delta_{(v_\kappa)})),\ v_{u,\gamma}(\phi \rightarrow \psi) = d]$.

What this means is that there are ordinals, the reflection ordinals, that capture all the stabilities in the sequence (v_κ). Following Field (2016) we can now extend, for the integral values, the previous theorem to all the sentences in $\mathcal{L}^+$. Before that, though, let us prove the following lemma.

Lemma 7.2.5 (Continuity lemma) *The value of the $\rightarrow$-conditionals at a point u is continuous at limit ordinals. That is, for any $u \in W$, any ϕ and ψ in $\mathcal{L}^+$ and any limit ordinal λ,*

$$v_{u,\lambda}(\phi \rightarrow \psi) = \begin{cases} 1 & \textit{iff } (\exists \beta < \lambda)(\forall \gamma \in [\beta, \lambda)), \quad v_{u,\gamma}(\phi \rightarrow \psi) = 1 \\ 0 & \textit{iff } (\exists \beta < \lambda)(\forall \gamma \in [\beta, \lambda)), \quad v_{u,\gamma}(\phi \rightarrow \psi) = 0 \\ 1/2 & \textit{otherwise.} \end{cases}$$

Proof From right to left. Suppose that there is a β, $\beta < \lambda$, such that for every γ, $\gamma \in [\beta, \lambda)$, $v_{u,\gamma}(\phi \rightarrow \psi) = 1$. In particular, then, there is a β, $\beta < \lambda$, such that for every γ, $\gamma \in [\beta, \lambda)$, $v_{u,\gamma+1}(\phi \rightarrow \psi) = 1$, which means that $|\phi|_{\langle u,j,\gamma\rangle} \leq |\psi|_{\langle u,j,\gamma\rangle}$. Consequently, $v_{u,\lambda}(\phi \rightarrow \psi) = 1$. Similarly for 0.

From left to right. The proof is by induction on the limit ordinal λ. Suppose that for each limit ordinal μ, $\mu < \lambda$, the result holds. Suppose, next, that $v_{u,\lambda}(\phi \rightarrow \psi) = 1$. Hence, there is a β, $\beta < \lambda$, such that for every γ, $\gamma \in [\beta, \lambda)$, $|\phi|_{\langle u,j,\gamma\rangle} \leq |\psi|_{\langle u,j,\gamma\rangle}$. Consequently, for every γ, $\gamma \in [\beta, \lambda)$, $v_{u,\gamma+1}(\phi \rightarrow \psi) = 1$. Finally, by induction hypothesis, for every γ, $\gamma \in [\beta + 1, \lambda)$, $v_{u,\gamma}(\phi \rightarrow \psi) = 1$. Similarly for 0. □

This lemma has important consequences for the semantics presented so far, as the following corollary, adapted from Field (2016), shows.

Corollary 7.2.6 *For any reflection ordinal $\Delta_{(v_\kappa)}$ for the sequence (v_κ), any $u \in W$, any transparent valuation j and any sentence $\phi \in \mathcal{L}^+$,*

(i) $|\phi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} = 1$ *if, and only if,* $\forall \theta \in COFIN_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} = 1$

(ii) $|\phi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} = 0$ *if, and only if,* $\forall \theta \in COFIN_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} = 0$

Proof The right to left direction of both (i) and (ii) is trivial, for $\Delta_{v_\kappa} \in \mathrm{COFIN}_{(v_\kappa)}$. So it remains to show their left to right direction.

Before proceeding with the proof, notice that, at first glance, it would be natural to adopt as strategy an induction on the complexity of ϕ. Unfortunately, this strategy will not work because we do not have any guarantee that, given a sentence of the form $Tr^{\ulcorner}\psi^{\urcorner}$, ψ will be a formula of a complexity less than $Tr^{\ulcorner}\psi^{\urcorner}$ itself. Our strategy will be to prove the contrapositives of (i) and (ii). But, since the problem is the predicate Tr and, consequently, we need to make explicit the stages σ of Kripke's fixed point construction[9],we will prove something stronger:

[9]See chapter 5.

(i*) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \neq 1$, then $\forall\sigma, |\phi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} \neq 1$

(ii*) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \neq 0$, then $\forall\sigma, |\phi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} \neq 0$

Notice that, properly, the consequents of (i*) and (ii*) are stronger than what we need to prove, for we need to conclude, from the supposition that $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$ such that $|\phi|_{\langle u,j,\theta\rangle} \neq 1$, that $|\phi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} \neq 1$. But given the fixed-point construction, if, for any σ, $|\phi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} \neq 1$ then ϕ will not have semantic value 1 at the corresponding fixed point $\boldsymbol{K}$, that is, $|\phi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} \neq 1$ (and the same for 0). So let us prove (i*) and (ii*) by induction on the stages σ of the fixed-point construction.

1. Base case. Suppose $\sigma = 0$. We need to show, then, the following:

   (i**) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \neq 1$, then $|\phi|_{\langle u,j,\Delta_{(v_\kappa)},0\rangle} \neq 1$

   (ii**) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \neq 0$, then $|\phi|_{\langle u,j,\Delta_{(v_\kappa)},0\rangle} \neq 0$

   This will be proven by induction on the complexity of ϕ.

   (a) First, when ϕ is an atomic formula of $\mathcal{L}$ or a $\Rightarrow$-conditional the result is clear, for their value do not change throughout the construction.

   (b) Next, suppose that $\phi = \neg\psi$ and that $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}, |\neg\psi|_{\langle u,j,\theta\rangle} \neq 1$, then $|\psi|_{\langle u,j,\theta\rangle} \neq 0$ and, by induction hypothesis, $|\psi|_{\langle u,j,\Delta_{(v_\kappa)},0\rangle} \neq 0$, which means that $|\neg\phi|_{\langle u,j,\Delta_{(v_\kappa)},0\rangle} \neq 1$. The result for the 0-clause for negation and the rest of the logical constants, except for $\rightarrow$, follow in a similar way.

   (c) Suppose, now, that $\phi = \psi \rightarrow \chi$. Since the value of $\rightarrow$-conditionals does not change throughout the fixed point construction and it depends only on the valuation v, what we need to prove is the following:

      (i') If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $v_{u,\theta}(\psi \rightarrow \chi) \neq 1$, then $v_{u,\Delta_{(v_\kappa)}}(\psi \rightarrow \chi) \neq 1$

      (ii') If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $v_{u,\theta}(\psi \rightarrow \chi) \neq 0$, then $v_{u,\Delta_{(v_\kappa)}}(\psi \rightarrow \chi) \neq 0$

      Given lemma 7.2.5 and the fact that $\Delta_{(v_\kappa)}$ is a limit ordinal, this amounts to the following:

      (i") If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $v_{u,\theta}(\psi \rightarrow \chi) \neq 1$, then $(\forall\beta < \Delta_{(v_\kappa)})(\exists\gamma \in [\beta, \Delta_{(v_\kappa)}))$, $v_{u,\gamma}(\psi \rightarrow \chi) \neq 1$

(ii") If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $v_{u,\theta}(\psi \to \chi) \neq 0$, then $(\forall\beta < \Delta_{(v_\kappa)})(\exists\gamma \in [\beta, \Delta_{(v_\kappa)}))$, $v_{u,\gamma}(\psi \to \chi) \neq 0$

Now, (i") and (ii") follow immediately by contraposition from the Reflection theorem 7.2.4.

(d) Finally, we must show that (i**) and (ii**) hold for $\phi = Tr^{\ulcorner}\psi^{\urcorner}$. But this is clear, for at the first stage of the fixed-point construction all Tr-ascriptions are assigned ½, so that the consequents of both (i**) and (ii**) are true.

2. Successor case. Suppose that the result holds for σ in order to show that it also holds for $\sigma + 1$. So we need to prove the following:

(i***) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \neq 1$, then $|\phi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma+1\rangle} \neq 1$

(ii***) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \neq 0$, then $|\phi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma+1\rangle} \neq 0$

Again, the proof proceeds by induction on the complexity of ϕ. The cases for atomic formulas of $\mathcal{L}$, $\Rightarrow$-conditionals, $\to$-conditionals and the rest of logical connectives are analogous as in the base case of the induction on σ. It remains to be proven that the result holds for sentences $\phi = Tr^{\ulcorner}\psi^{\urcorner}$.

Take the induction hypothesis on σ applied to ψ. Given that all v_κ are transparent and j is transparent we can apply proposition 7.2.3 and replace ψ by $Tr^{\ulcorner}\psi^{\urcorner}$ in the antecedents of (i***) and (ii***):

(i****) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,\theta\rangle} \neq 1$, then $|\psi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} \neq 1$

(ii****) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,\theta\rangle} \neq 0$, then $|\psi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} \neq 0$

Finally, notice that, given the fixed point construction, $|\psi|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} = |Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,\Delta_{(v_\kappa)},\sigma+1\rangle}$, which gives the desired result.

3. Limit case. Suppose λ is a limit ordinal and suppose that the result holds for all ordinals σ, $\sigma < \lambda$. As before, this is proven by induction on the complexity of ϕ and all the cases are analogous to the base case for σ except for sentences $\phi = Tr^{\ulcorner}\psi^{\urcorner}$. Hence, we need to show the following:

(i*) If $\exists\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,\theta\rangle} \neq 1$, then $|Tr^{\ulcorner}\psi^{\urcorner}|_{\langle u,j,\Delta_{(v_\kappa)},\lambda\rangle} \neq 1$

(ii*) If $\exists\theta \in \text{COFIN}_{(v_\kappa)}$, $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\theta\rangle} \neq 0$, then $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\Delta_{(v_\kappa)},\lambda\rangle} \neq 0$

To see that (i*) is the case, suppose that there is a $\theta \in \text{COFIN}_{(v_\kappa)}$ such that $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\theta\rangle} \neq 1$ and that $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\Delta_{(v_\kappa)},\lambda\rangle} = 1$. Then, by the induction hypothesis, for all σ, $\sigma < \lambda$, $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} \neq 1$. On the other hand, if $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\Delta_{(v_\kappa)},\lambda\rangle} = 1$, then, given the fixed-point construction, there has to be a σ, $\sigma < \lambda$, $|Tr^\ulcorner\psi^\urcorner|_{\langle u,j,\Delta_{(v_\kappa)},\sigma\rangle} = 1$. Contradiction. Similarly for 0.□

We need now to take stock. For the moment, ignore the valuation j and think of the above result as applied to a language $\mathcal{L}'^+$ like $\mathcal{L}^+$ but without $\Rightarrow$. Then, this construction tell us how to define a privileged valuation, $v_{\Delta_{(v_\kappa)}}$, for any reflection ordinal $\Delta_{(v_\kappa)}$, that governs the values of the $\rightarrow$-conditionals. Now we can define validity in terms of it. Thus, for any sentence δ of $\mathcal{L}'^+$ and any set of sentences Γ of $\mathcal{L}'^+$ we say that $\Gamma \models \delta$ if, and only if, for all models (we are ignoring now W so that there is no need to mention the points), if $|\Gamma|_{\langle\Delta_{(v_\kappa)}\rangle} = 1$ then $|\delta|_{\langle\Delta_{(v_\kappa)}\rangle} = 1$.[10]

With these definitions at hand we obtain one of the logics defended in Field (2003a, 2008).[11] The main problem of this conditional is that it still seems too weak. Some principles we would like to have are not satisfied in the logic. For example, these axioms of *RWK* as presented above are not among the principles of $\rightarrow$:

5. $((\phi \rightarrow \psi) \wedge (\phi \rightarrow \chi)) \rightarrow (\phi \rightarrow (\psi \wedge \chi))$

12. $\phi \rightarrow (\psi \rightarrow \phi)$

Field (2014, 2016) also considers $\rightarrow$ to be too weak (see, for example Field 2016, p. 1), for it cannot adequately express restricted quantification; that is, we cannot express principles like, for example, the following ones:

$$(\forall x(\phi x \rightarrow \psi x) \wedge \forall x(\phi x \rightarrow \chi x)) \rightarrow \forall x(\phi x \rightarrow (\psi x \wedge \chi x))$$

[10] Recall that $|\delta|_{\langle\Delta_{(v_\kappa)}\rangle}$ is an abbreviation of $|\delta|_{\langle\Delta_{(v_\kappa)},\boldsymbol{K}\rangle}$, where $\boldsymbol{K}$ is the minimal fixed point of Kripke's construction with $v_{\Delta_{(v_\kappa)}}$ as the valuation for $\rightarrow$-conditionals.

[11] The main logic defended in Field (2003a, 2008) has a conditional that is slightly different in the 0-clause; its 0-clause is satisfied when the condition that stabilizes in the sequence is that the value of the antecedent is strictly greater than the value of the consequent. The conditional $\rightarrow$ is called by Field (2008) *the first variant*. Filed himself adopts the first variant in Field (2014, 2016).

$$\forall x \phi x \rightarrow \forall x(\psi x \rightarrow \phi x)$$

which cannot be obtained unless 5 and 12 above hold. Field (2014, 2016) has tried to overcome this difficulty by adding a new conditional based on the sort of so called *variably strict conditionals* discussed in Stalnaker (1968) or Lewis (1974), among many others. Field (2014, 2016) uses two conditionals, $\blacktriangleright$ and $\triangleright$. The former is the conditional he used in Field (2008), $\rightarrow$ in this chapter, which is, in Field's words, *material-like* . The latter is a variably strict conditional defined intensionally over a set of points intended to capture indicative conditionals. Field can obtain, then, among others, mixed versions of the principles above:

5. $((\phi \blacktriangleright \psi) \wedge (\phi \blacktriangleright \chi)) \triangleright (\phi \blacktriangleright (\psi \wedge \chi))$

12. $\phi \triangleright (\psi \blacktriangleright \phi)$

Hence, Field uses a material-like conditional to express the conditional used to restrict universal quantification and another conditional used to capture ordinary uses of the indicative conditional.

I am not concerned about restricted quantification here, but about how we can strengthen the logic in a language with a truth predicate and vague predicates, while preserving the semantics of the latter ones. Accordingly, I propose to apply the techniques in Field (2016) to add to the construction of $\rightarrow$ a new conditional, $\Rightarrow$, intended to capture certain analytic relations that occur between sentences of $\mathcal{L}^{+}$. In particular, I want to see whether $\Rightarrow$ can capture the relation of entailment and the penumbral connections. In the next sections we will see how this conditional works and which principles can we have once we define validity, so that the new conditional satisfies laws we would like to have in the logic. Next, we will see whether penumbral connections can be captured. All of this, of course, with a truth predicate, Tr, satisfying the Intersubstitutivity Principle introduced at page 6.

7.2 The Second Conditional

Recall that we left a valuation j for $\Rightarrow$-conditionals fixed and, over it, we built a privileged valuation for $\rightarrow$-conditionals, let us call it v, that depended on the model for $\mathcal{L}$ and j. We also saw that v was transparent and, hence, given proposition 7.2.3, if we obtain a transparent j we will retain intersubstitutivity for Tr. The strategy I will follow will be to add the conditional $\Rightarrow$ to the construction by using the techniques in Field (2016), which consist in devising another revision sequence for j, (j_{κ}), that is defined over the class of all ordinals and that eventually yields the intended valuation

for $\Rightarrow$-conditionals. I will write $|\phi|_{\langle u,\kappa\rangle}$ instead of $|\phi|_{\langle u,j_\kappa,\Delta_{(v_\kappa)},\boldsymbol{K}\rangle}$, so that the former is the semantic value of the sentence ϕ under the fixed point $\boldsymbol{K}$ that we obtain when we evaluate $\rightarrow$-conditionals under $v_{\Delta_{(v_\kappa)}}$ (where $\Delta_{(v_\kappa)}$ is any reflection ordinal for (v_κ)) using j_κ as the valuation for $\Rightarrow$-conditionals.

As before, let us begin with a function j_0 that assigns $1/2$ to all the sentences of the form $\phi \Rightarrow \psi$. The rule is the following one, for each $u \in W$:

$$j_{u,\alpha}(\phi \Rightarrow \psi) = \begin{cases} 1 \text{ iff } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha))(\forall w \in W, u \leq w), \\ \text{ if } |\phi|_{\langle w,\gamma\rangle} = 1, \text{ then } |\psi|_{\langle w,\gamma\rangle} = 1 \\ 0 \text{ iff } (\exists\beta < \alpha)(\forall\gamma \in [\beta,\alpha))(\forall w \in W, u \leq w)(\exists v \in W, w \leq v), \\ |\phi|_{\langle v,\gamma\rangle} = 1 \text{ and } |\psi|_{\langle v,\gamma\rangle} = 0 \\ 1/2 \text{ otherwise.} \end{cases}$$

As before, the reflection theorem 7.2.4 and the Continuity lemma 7.2.5 hold for (j_κ). Unfortunately, though, corollary 7.2.6 does not hold unrestrictedly. What this corollary would claim, applied to (j_κ), is that for any reflection ordinal $\Omega_{(j_\kappa)}$, any $u \in W$ and any sentence $\phi \in L^+$, the valuation $j_{\Omega_{(j_\kappa)}}$ captures the behavior of ϕ in the series. But this is not the case for $\rightarrow$-conditionals. To see this, consider the following example from Field (2016). Let us have a sentence, $\lambda_\Rightarrow$, that is equivalent to $Tr^\ulcorner\lambda_\Rightarrow{}^\urcorner \Rightarrow \neg Tr^\ulcorner\lambda_\Rightarrow{}^\urcorner$, which, given intersubstitutivity of truth, is equivalent to $\lambda_\Rightarrow \Rightarrow \neg\lambda_\Rightarrow$. Notice that at each stage κ of the (j_κ) series and any $u \in W$,

(a) if κ is a limit ordinal, then $|\lambda_\Rightarrow|_{\langle u,\kappa\rangle} = 1/2$,

(b) if κ is an odd successor, then $|\lambda_\Rightarrow|_{\langle u,\kappa\rangle} = 1$,

(c) if κ is an even successor, then $|\lambda_\Rightarrow|_{\langle u,\kappa\rangle} = 0$.

Next, consider the sentence $\lambda_\rightarrow$ equivalent to $Tr^\ulcorner\lambda_\Rightarrow{}^\urcorner \rightarrow \neg Tr^\ulcorner\lambda_\Rightarrow{}^\urcorner$ and, hence, equivalent to $\lambda_\Rightarrow \rightarrow \neg\lambda_\Rightarrow$. Since $\lambda_\Rightarrow$ is equivalent to a $\Rightarrow$-conditional, its value, given by a valuation j_κ, does not change through the revision sequence for each v_μ over (j_κ). That means that, for any $u \in W$ and any ordinals κ and $\mu > 0$,

(d) if $|\lambda_\Rightarrow|_{\langle u,\kappa\rangle} \neq 1$, then $|\lambda_\rightarrow|_{\langle u,\kappa,\mu\rangle} = 1$,

(e) if $|\lambda_\Rightarrow|_{\langle u,\kappa\rangle} = 1$, then $|\lambda_\rightarrow|_{\langle u,\kappa,\mu\rangle} = 0$.

Combining (a)-(e) we get, for any ordinals κ and $\mu > 0$,

(f) if κ is an even successor or a limit, then $|\lambda_{\rightarrow}|_{\langle u,\kappa,\mu\rangle} = 1$,

(g) if κ is an odd successor, then $|\lambda_{\rightarrow}|_{\langle u,\kappa,\mu\rangle} = 0$.

But, since all reflection ordinals $\Omega_{(j_\kappa)}$ are limit ordinals, then, for any $u \in W$, $|\lambda_{\rightarrow}|_{\langle u,\Omega_{(j_\kappa)}\rangle} = 1$, in spite of the fact that the value of $\lambda_{\rightarrow}$ is not 1 at every cofinal ordinal. Thus, the result of corollary 7.2.6 does not hold any more for $\rightarrow$-conditionals. Still, it does hold for $\Rightarrow$-conditionals so that we have a restricted version of it.

Corollary 7.2.7 (Field 2016) *For any reflection ordinal $\Omega_{(j_\kappa)}$ for the sequence (j_κ), any $u \in W$ and any sentence in $\mathcal{L}^+$ of the form $\phi \Rightarrow \psi$,*

(i) $|\phi \Rightarrow \psi|_{\langle u,\Omega_{(j_\kappa)}\rangle} = 1$ *if, and only if,* $\forall\theta \in COFIN_{(j_\kappa)}$, $|\phi \Rightarrow \psi|_{\langle u,\theta\rangle} = 1$

(ii) $|\phi \Rightarrow \psi|_{\langle u,\Omega_{(j_\kappa)}\rangle} = 0$ *if, and only if,* $\forall\theta \in COFIN_{(j_\kappa)}$, $|\phi \Rightarrow \psi|_{\langle u,\theta\rangle} = 0$

The proof is analogous to the part of the corollary 7.2.6 that deals with $\rightarrow$.

Although the limitation in corollary 7.2.7 represents a difficulty, we can still try to use the reflection ordinals for the sequence (j_κ) to define validity and see how the conditional $\Rightarrow$ behaves. Moreover, since, again, j_0 was clearly transparent and the rule for (j_κ) preserves transparency, proposition 7.2.3 guarantees intersubstitutivity for Tr in the whole construction, so that for any $u \in W$, any ordinal κ and any sentence $\phi \in \mathcal{L}$, $|\phi|_{\langle u,\kappa\rangle} = |Tr^{\ulcorner}\phi^{\urcorner}|_{\langle u,\kappa\rangle}$.

Let us see what has been done so far. The semantic value of a sentence depends on four parameters. First, on a point $u \in W$ in the model for $\mathcal{L}$ that only affects the conditional $\Rightarrow$. Second, an ordinal that represents the stage in the sequence of valuations for $\Rightarrow$-conditionals (j_κ). We arrived at the intended valuation in this series which was $j_{\Omega_{(j_\kappa)}}$. Third, an ordinal that represents the stage in the sequence for $\rightarrow$-conditionals (v_κ) that depended on our choice of valuation j. Again, we arrived at the intended valuation in this series which was $v_{\Delta_{(v_\kappa)}}$. Fourth, an ordinal that represents the stage in the Kripke's fixed-point construction and that depends on our choice of valuation j and v. In symbols, this would be $|\phi|_{\langle u,\Omega_{(j_\kappa)},\Delta_{(v_\kappa)},\boldsymbol{K}\rangle}$, which I abbreviate when it is possible.

7.2 Validity

We can now define validity in terms of the reflection ordinals for (j_κ). We say that a sentence γ is a logical consequence of a set of sentences Γ, in symbols

$\Gamma \models \gamma$, when for every model $\mathcal{M}$, every $u \in W^{\mathcal{M}}$, any reflection ordinal for the sequence (j_κ), $\Omega_{(j_\kappa)}$, and any reflection ordinal for the sequence (v_κ) over $\Omega_{(j_\kappa)}$, $\Delta_{(v_\kappa)}$, if $|\Gamma|_{\langle u,\Omega_{(j_\kappa)},\Delta_{(v_\kappa)}\rangle} = 1$, then $|\gamma|_{\langle u,\Omega_{(j_\kappa)},\Delta_{(v_\kappa)}\rangle} = 1$. As usual, validity is defined as logical consequence from the empty set. As before, I will sometimes simplify, for readability, $|\gamma|_{\langle u,\Omega_{(j_\kappa)},\Delta_{(v_\kappa)}\rangle}$ with $|\gamma|_{\langle u,\Omega_{(j_\kappa)}\rangle}$.

Notice that, given transparency of j and v, proposition 7.2.3 and the definition of logical consequence, Tr satisfies the Intersubstitutivity Principle.

It will be useful to have the following lemma.

Lemma 7.2.8 (Field 2016) *For any valuation j for $\Rightarrow$-conditionals, any reflection ordinal $\Delta_{(v_\kappa)}$ for (v_κ) over j, any $u \in W$ and any $\rightarrow$-conditional $\phi \rightarrow \psi$, $|\phi \rightarrow \psi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} = 1$ if, and only if, for all $\theta \in COFIN_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \leq |\psi|_{\langle u,j,\theta\rangle}$.*

Proof Take any valuation j for $\Rightarrow$-conditionals, any reflection ordinal $\Delta_{(v_\kappa)}$ for v_κ over j, any $u \in W$ and any $\rightarrow$-conditional $\phi \rightarrow \psi$. For the left to right direction suppose that $|\phi \rightarrow \psi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} = 1$. Then, by corollary 7.2.6, for each $\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi \rightarrow \psi|_{\langle u,j,\theta\rangle} = 1$. Hence, for each $\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi \rightarrow \psi|_{\langle u,j,\theta+1\rangle} = 1$, which, given the definition of $\rightarrow$, implies the desired result.

For the right to left part suppose that for all $\theta \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle u,j,\theta\rangle} \leq |\psi|_{\langle u,j,\theta\rangle} = 1$. Then, $|\phi \rightarrow \psi|_{\langle u,j,\theta+1\rangle} = 1$ which, given the Continuity lemma 7.2.5, implies that there is an ordinal α such that, for all ordinals β, $\alpha \leq \beta$, $|\phi \rightarrow \psi|_{\langle u,j,\beta\rangle} = 1$. Finally, since $\Delta_{(v_\kappa)}$ is a cofinal ordinal, $|\phi \rightarrow \psi|_{\langle u,j,\Delta_{(v_\kappa)}\rangle} = 1$. □

This lemma also holds for $\Rightarrow$-conditionals and reflection ordinals $\Omega_{(j_\kappa)}$:

Lemma 7.2.9 *For any $u \in W$, any reflection ordinal $\Omega_{(j_\kappa)}$ for (j_κ), any reflection ordinal for the sequence (v_κ) over $\Omega_{(j_\kappa)}$, $\Delta_{(v_\kappa)}$, and any $\Rightarrow$-conditional $\phi \Rightarrow \psi$, $|\phi \Rightarrow \psi|_{\langle u,\Omega_{(j_\kappa)},\Delta_{(v_\kappa)}\rangle} = 1$ if, and only if, for any $w \in W$, $u \leq w$, any ordinal $\theta \in COFIN_{(j_\kappa)}$ and any reflection ordinal for the sequence (v_κ) over θ, $\Delta^{\theta}_{(v_\kappa)}$, if $|\phi|_{\langle w,\theta,\Delta^{\theta}_{(v_\kappa)}\rangle} = 1$ then $|\psi|_{\langle w,\theta,\Delta^{\theta}_{(v_\kappa)}\rangle} = 1$.*

Proof The proof is essentially the same as in lemma 7.2.8.

The following schemas all have valid instances in the construction above:

1. $\phi \Rightarrow \phi$
2. $\phi \Rightarrow (\phi \vee \psi)$

3. $(\phi \wedge \psi) \Rightarrow \phi$
4. $(\phi \wedge (\psi \vee \chi)) \Rightarrow ((\phi \wedge \psi) \vee (\phi \wedge \chi))$
5. $((\phi \rightarrow \psi) \wedge (\phi \rightarrow \chi)) \Rightarrow (\phi \rightarrow (\psi \wedge \chi))$
6. $((\phi \rightarrow \chi) \wedge (\psi \rightarrow \chi)) \Rightarrow ((\phi \vee \psi) \rightarrow \chi)$
7. $\neg\neg\phi \Rightarrow \phi$
8. $(\phi \rightarrow \neg\psi) \Rightarrow (\psi \rightarrow \neg\phi)$
9. $(\phi \rightarrow \psi) \Rightarrow ((\psi \rightarrow \chi) \rightarrow (\phi \rightarrow \chi))$
10. $(\phi \rightarrow \psi) \Rightarrow ((\chi \rightarrow \phi) \rightarrow (\chi \rightarrow \psi))$
11. $(\phi \wedge (\phi \rightarrow \psi)) \Rightarrow \psi$
12. $\phi \Rightarrow (\psi \rightarrow \phi)$
13. $\forall x\phi \Rightarrow \phi(d/x)$
14. $\forall x(\phi \rightarrow \psi) \Rightarrow (\phi \rightarrow \forall x\psi)$
15. $\forall x(\psi \rightarrow \phi) \Rightarrow (\exists x\psi \rightarrow \phi)$
16. $\forall x(\phi \vee \psi) \Rightarrow (\phi \vee \forall x\psi)$[12]

Proof Principles 1 and 7 are clear, for the antecedent and the consequent will have the same truth value at every $u \in W$ and every reflection ordinal $\Omega_{(j_\kappa)}$.

Principle 2. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|\phi \Rightarrow (\psi \vee \phi)|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. By lemma 7.2.9, we conclude that there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W$, $u \leq w$, such that $|\phi|_{\langle w,\theta\rangle} = 1$ and $|\phi \vee \psi|_{\langle w,\theta\rangle} \neq 1$, which, given the valuation for $\vee$, yields a contradiction. Principles 3 and 4 are similar; all follow from the fact that the value of the antecedent will always be less than or equal to the value of the consequent. Principle 13, which involves the universal quantifier, is also similar.

Principle 5. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|((\phi \rightarrow \psi) \wedge (\phi \rightarrow \chi)) \Rightarrow (\phi \rightarrow (\psi \wedge \chi))|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. Then, by lemma 7.2.9, there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W^{\mathcal{M}}$, $u \leq w$, such

[12] In 14, 15 and 16 x is not free in ϕ.

that, for some reflection ordinal $\Delta_{(v_\kappa)}$ for (v_κ) over j_θ, (i) $|(\phi \to \psi) \wedge (\phi \to \chi)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and (ii) $|\phi \to (\psi \wedge \chi)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} \neq 1$.

By (i) we get that $|\phi \to \psi|_{w,\theta,\Delta_{(v_\kappa)}} = 1$ and $|\phi \to \chi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$, which, by lemma 7.2.8, imply that for all $\rho \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle w,\theta,\rho\rangle} \leq |\psi|_{\langle w,\theta,\rho\rangle}$ and $|\phi|_{\langle w,\theta,\rho\rangle} \leq |\chi|_{\langle w,\theta,\rho\rangle}$.

By (ii) and lemma 7.2.8, we conclude that there is a $\tau \in \mathrm{COFIN}_{(v_\kappa)}$ such that $|\phi|_{\langle w,\theta,\tau\rangle} > |\psi \wedge \chi|_{\langle w,\theta,\tau\rangle}$. Contradiction. Principle 6 is similar.

Principle 8. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|(\phi \to \neg\psi) \Rightarrow (\psi \to \neg\phi)|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. By lemma 7.2.9, we conclude that there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W$, $u \leq w$, such that (i) $|\phi \to \neg\psi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and (ii) $|\psi \to \neg\phi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} \neq 1$.

By (i) we get that there is a β, $\beta < \Delta_{(v_\kappa)}$, such that for all γ, $\gamma \in [\beta, \Delta_{(v_\kappa)})$, $|\phi|_{\langle w,\theta,\gamma\rangle} \leq |\neg\psi|_{\langle w,\theta,\gamma\rangle}$. And by (ii), we obtain that for all β, $\beta < \Delta_{(v_\kappa)}$, there is a γ, $\gamma \in [\beta, \Delta_{(v_\kappa)})$ such that $|\psi|_{\langle w,\theta,\gamma\rangle} > |\neg\phi|_{\langle w,\theta,\gamma\rangle}$. This means that there is a τ, $\tau < \Delta_{(v_\kappa)}$ such that $|\phi|_{\langle w,\theta,\tau\rangle} \leq |\neg\psi|_{\langle w,\theta,\tau\rangle}$ and $|\psi|_{\langle w,\theta,\tau\rangle} > |\neg\phi|_{\langle w,\theta,\tau\rangle}$, which, given the definition of $\neg$, is impossible.

Principle 10. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|(\phi \to \psi) \Rightarrow ((\chi \to \phi) \to (\chi \to \psi))|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. Then, by lemma 7.2.9, there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W^{\mathcal{M}}$, $u \leq w$, such that (i) $|\phi \to \psi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and (ii) $|(\chi \to \phi) \to (\chi \to \psi)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} \neq 1$.

By (i) and lemma 7.2.8, we get that for all $\rho \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle w,\theta,\rho\rangle} \leq |\psi|_{\langle w,\theta,\rho\rangle}$.

By (ii) and lemma 7.2.8, we get that there is a $\tau \in \mathrm{COFIN}_{(v_\kappa)}$ such that $|\chi \to \phi|_{\langle w,\theta,\tau\rangle} > |\chi \to \psi|_{\langle w,\theta,\tau\rangle}$. Since τ is cofinal, there will be another cofinal ordinal $\tau' > \tau$, such that $v_\tau = v_{\tau'}$ and, hence, such that $|\chi \to \phi|_{\langle w,\theta,\tau'\rangle} > |\chi \to \psi|_{\langle w,\theta,\tau'\rangle}$. Now, we have two disjunctive options both of which are contradictory,

- First, $|\chi \to \phi|_{\langle w,\theta,\tau'\rangle} = 1$ and $|\chi \to \psi|_{\langle w,\theta,\tau'\rangle} \neq 1$. The first conjunct implies, by the definition of $\to$, that there is a $\mu < \tau'$, such that, for all $\gamma \in [\mu, \tau')$, $|\chi|_{\langle w,\theta,\gamma\rangle} \leq |\phi|_{\langle w,\theta,\gamma\rangle}$ and the second conjunct implies that for all $\alpha < \tau'$, there is a $\beta \in [\alpha, \tau')$, $|\chi|_{\langle w,\theta,\beta\rangle} > |\psi|_{\langle w,\theta,\beta\rangle}$. Consider now $\zeta = \sup\{\mu, \tau\}$. Since τ is cofinal, so is ζ and, moreover, since $\zeta < \tau'$, there is a $\pi \in [\zeta, \tau')$ such that $|\chi|_{\langle w,\theta,\pi\rangle} \leq |\phi|_{\langle w,\theta,\pi\rangle}$ and $|\chi|_{\langle w,\theta,\pi\rangle} > |\psi|_{\langle w,\theta,\pi\rangle}$. Since π is also cofinal, by (i) we get $|\phi|_{\langle w,\theta,\pi\rangle} \leq |\psi|_{\langle w,\theta,\pi\rangle}$, which is impossible.

- Second, $|\chi \to \psi|_{\langle w,\theta,\tau'\rangle} = 0$ and $|\chi \to \phi|_{\langle w,\theta,\tau'\rangle} \neq 0$. In this case, there is a $\mu < \tau'$, such that, for all γ, $\gamma \in [\mu, \tau')$, $|\chi|_{\langle w,\theta,\gamma\rangle} = 1$ and $|\psi|_{\langle w,\theta,\gamma\rangle} = 0$, and for all $\alpha < \tau'$, there is a $\beta \in [\alpha, \tau')$, such that either $|\chi|_{\langle w,\theta,\beta\rangle} \neq 1$

or $|\phi|_{\langle w,\theta,\beta\rangle} \neq 0$. As in the previous case, this implies that there is a cofinal π such that $|\chi|_{\langle w,\theta,\pi\rangle} = 1$ and $|\psi|_{\langle w,\theta,\pi\rangle} = 0$ and $|\phi|_{\langle w,\theta,\pi\rangle} \neq 0$. Moreover, since π is cofinal, by (i) we get $|\phi|_{\langle w,\theta,\pi\rangle} \leq |\psi|_{\langle w,\theta,\pi\rangle}$, which is impossible.

Principle 9 is similar.

Principle 11. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|(\phi \wedge (\phi \rightarrow \psi)) \Rightarrow \psi)|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. Then there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W$, $u \leq w$, such that (i) $|\phi \wedge (\phi \rightarrow \psi)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and (ii) $|\psi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} \neq 1$.

By (i) we get that $|\phi|_{w,\theta,\Delta_{(v_\kappa)}} = 1$ and $|\phi \rightarrow \psi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$, which, by lemma 7.2.8 and corollary 7.2.7, imply that for all $\rho \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle w,\theta,\rho\rangle} = 1$ and $|\phi|_{\langle w,\theta,\rho\rangle} \leq |\psi|_{\langle w,\theta,\rho\rangle}$. But this means that for all $\rho \in \mathrm{COFIN}_{(v_\kappa)}$, $|\psi|_{\langle w,\theta,\rho\rangle} = 1$, contradicting (ii).

Principle 12. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|\phi \Rightarrow (\psi \rightarrow \phi)|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. Then, there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W^{\mathcal{M}}$, $u \leq w$, such that (i) $|\phi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and (ii) $|\psi \rightarrow \phi|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} \neq 1$.

By (i) and corollary 7.2.6, we get that for all $\rho \in \mathrm{COFIN}_{(v_\kappa)}$, $|\phi|_{\langle w,\theta,\rho\rangle} = 1$. Moreover, (ii) together with lemma 7.2.8, imply that there is a $\tau \in \mathrm{COFIN}_{(v_\kappa)}$ such that $|\psi|_{\langle w,\theta,\tau\rangle} > |\phi|_{\langle w,\theta,\tau\rangle}$. But, since by the previous considerations, $|\phi|_{\langle w,\theta,\tau\rangle} = 1$, we reach a contradiction.

Principle 14. Suppose that there is a model $\mathcal{M}$, a $u \in W^{\mathcal{M}}$ and a reflection ordinal $\Omega_{(j_\kappa)}$ such that $|\forall x(\phi \rightarrow \psi) \Rightarrow (\phi \rightarrow \forall x\psi)|_{\langle u,\Omega_{(j_\kappa)}\rangle} \neq 1$. Then there is a $\theta \in \mathrm{COFIN}_{(j_\kappa)}$ and a $w \in W$, $u \leq w$, such that (i) $|\forall x(\phi \rightarrow \psi)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and (ii) $|(\phi \rightarrow \forall x\psi)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} \neq 1$.

By (i), $\min\{|(\phi \rightarrow \psi)(d/x)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} : d \in D\} = 1$. Consequently, for each $d \in D$, $|(\phi \rightarrow \psi)(d/x)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$ and, since x is not free in ϕ, $|\phi \rightarrow \psi(d/x)|_{\langle w,\theta,\Delta_{(v_\kappa)}\rangle} = 1$. Hence, for each d, there is a β_d, $\beta_d < \Delta_{(v_\kappa)}$ such that, for all γ, $\gamma \in [\beta_d, \Delta_{(v_\kappa)})$, $|\phi|_{\langle w,\theta,\gamma\rangle} \leq |\psi(d/x)|_{\langle w,\theta,\gamma\rangle}$.

By (ii) and the definition of $\rightarrow$, we get that for all $\beta < \Delta_{(v_\kappa)}$ there is a $\gamma \in [\beta, \Delta_{(v_\kappa)})$, $|\phi|_{\langle w,\theta,\gamma\rangle} > |\forall x\psi|_{\langle w,\theta,\gamma\rangle} = \min\{|(\psi)(d/x)|_{\langle w,\theta,\gamma\rangle} : d \in D\}$. Therefore, we can choose a d_0 such that $|\phi|_{\langle w,\theta,\gamma\rangle} > |\psi(d_0/x)|_{\langle w,\theta,\gamma\rangle}$. Consider, next, β_{d_0}. Since $\beta_{d_0} < \Delta_{(v_\kappa)}$, then there is a $\gamma_0 \in [\beta_{d_0}, \Delta_{(v_\kappa)})$, such that $|\phi|_{\langle w,\theta,\gamma_0\rangle} > |\psi(d_0/x)|_{\langle w,\theta,\gamma_0\rangle}$. Moreover, by (i) and the fact that $\gamma_0 \in [\beta_{d_0}, \Delta_{(v_\kappa)})$, we obtain that $|\phi|_{\langle w,\theta,\gamma_0\rangle} \leq |\psi(d_0/x)|_{\langle w,\theta,\gamma_0\rangle}$, which is impossible. Principles 15 and 16 are similar. □

The behavior of $\rightarrow$ and $\Rightarrow$ alone are very similar. One difference, that

stems directly from the way they are defined, is that $\rightarrow$ is contrapositive while $\Rightarrow$ is not. Moreover, $\rightarrow$ does not satisfy principles 5, 6, 9, 10, 11, 12, 14, 15 and 16 (although all of them, except 12, are obtained in rule forms).[13]

As can be seen, in the mixed version, most of the principles of CK (all of them except 11) are satisfied in mixed form. If we take CK as capturing the logical laws we might want to have in a paracomplete logic with a vague predicate, we can then take $\Rightarrow$ as capturing the relation of entailment between sentences of $\mathcal{L}^{+}$.

Although the construction has one form of the *Modus Ponens* reasoning, principle 11, it lacks the following:

11'. $\phi \Rightarrow ((\phi \rightarrow \psi) \rightarrow \psi)$

To see why, consider again sentence $\lambda_{\Rightarrow}$. Since the value of $\lambda_{\Rightarrow}$ keeps oscillating from 1 to 0 along the successor ordinals of the (j_κ) series, the same happens eventually to the sentence $\top \Rightarrow ((\top \rightarrow \lambda_{\Rightarrow}) \rightarrow \lambda_{\Rightarrow})$, which means that it gets value ½ at all limit ordinals in the j_κ sequence, hence, in particular, it has semantic value ½ at all reflection ordinals $\Omega_{(j_\kappa)}$.

7.3 Vagueness

I want to sketch, next, one way of adapting the previous semantics to vagueness due to Field (2003c) and to propose another one that seems slightly more natural to me. As far as I can see, a kind of irresolvable tension arises at this point within what I have been calling *the paracomplete project*. The tension has to do with the role that the model is taken to be playing.

Field (2003c) presents a generalization of the semantics for $\rightarrow$ intended to be "the unified logic for vagueness and the [truth] paradoxes" (Field 2003c, p. 294). Moreover, he claims, the proposed semantics should be something more than a mere tool to give an extensionally adequate notion of logical consequence; it should represent the semantics of vague terms "as faithfully as possible" (Field 2003c, p. 289).

But, if we consider the constructions we used in this chapter to strengthen the conditional of SK, they can hardly be considered as faithfully representing the semantics of 'true'; indeed, this seems clear if we compare them with Kripke's construction, which, as we saw in chapter 5, can be seen as capturing some of the core intuitions behind the use of 'true'. As a matter of fact, the main purpose of the more sophisticated constructions considered in this

[13] The details about why these principles fail can be found in Field (2008, pp. 266–270).

chapter is to show how a truth predicate can be consistently added to any base model with a suitable logic.[14] What that means is that the intuitive appeal of Kripke's construction is abandoned.

Still, we can try to capture the semantics of vague predicates as faithfully as possible. This implies, at the very least, respecting the penumbral intuitions we discussed in chapter 5. So if a and b are borderline cases of the predicate 'tall' (T), then, if b is taller than a, we would like to assert the sentence $Ta \Rightarrow Tb$. Moreover, as some authors have claimed (like, for instance, Field 2003c or Shapiro 2006), given a borderline case of 'tall' and of 'short' (S), we should be able to say that a is in the short-to-tall region. But, given the semantics of $\vee$, we cannot express that with $Sa \vee Ta$. We can use the conditional, though, and try to express that a is in the short-to-tall region with $\neg Sa \Rightarrow Ta$. The same happens with the idea that 'short' and 'tall' are contraries, which, although it cannot be expressed with the use of $\vee$, it can be expressed with $\forall x(Sx \Rightarrow \neg Tx)$. So, taking stock, we need the following sentences to be true, for any a and b borderline case of 'tall' and 'short':

(i) $Ta \Rightarrow Tb$, when b is taller than a.

(ii) $\neg Sa \Rightarrow Ta$.

(iii) $\forall x(Sx \Rightarrow \neg Tx)$

Let me sketch, next, the proposal in Field (2003c). The semantics uses an infinite set W of worlds at which sentences are assigned one member of $\{0, 1/2, 1\}$ and a privileged world @. We must think of the elements of W as "alternative methods for assigning semantic values to actual and possible sentences, given the way the world actually is in precise respects" and @ as "the actual assignment" (Field 2003c, p. 290). Each $w \in W$ is assigned to a (possibly empty) *directed* family $\mathcal{F}_w$ of nonempty subsets of W, called *w-neighborhoods*. A given $\mathcal{F}_w$ is *directed* if, and only if,

$$(\forall \mathcal{X}, \mathcal{Y} \in \mathcal{F}_w)(\exists \mathcal{Z} \in \mathcal{F}_w), \mathcal{Z} \subseteq \mathcal{X} \cap \mathcal{Y}$$

Each w-neighborhood is meant to represent the worlds that meet a certain standard of similarity to w. Field proceeds by adding some conditions to @:

[14] As Field stresses in many places, these constructions show something more, a kind of conservativeness result; they show that a truth predicate can be added to a given base model without changing the true-free part of the language that depended on that base model (see, for instance, Field 2008, p. 66 or Field 2003a, p. 171).

1. @ is *normal*, that is, $(\forall \mathcal{X} \in \mathcal{F}_@), @ \in \mathcal{X}$
2. $\{@\} \notin \mathcal{F}_@$

That is, the actual assignment @ is similar to itself according to all standards of similarity and it shares any standard of similarity with at least another world. Field imposes other conditions to the worlds in W and their neighborhoods, but since they are not important for the purposes of this section, I will skip over them. As expected, a model will consist of a given domain for each $w \in W$ and an assignment of an extension and an anti-extension (as usual, two disjoint but not necessarily exhaustive subsets of the domain) to each predicate, so that vague predicates can be captured in the model. The valuation rules for the conditional-free fragment of the language are governed by *SK* with no reference to other worlds. The conditional $\rightarrow$ is defined as follows:

$$|\phi \rightarrow \psi|_u = \begin{cases} 1 & \text{iff } (\exists \mathcal{X} \in \mathcal{F}_u)(\forall w \in \mathcal{X}), |\phi|_w \leq |\psi|_w \\ 0 & \text{iff } (\exists \mathcal{X} \in \mathcal{F}_u)(\forall w \in \mathcal{X}), |\phi|_w = 1 \text{ and } |\psi|_w = 0 \\ \frac{1}{2} & \text{otherwise.} \end{cases}$$

Validity is defined in terms of preservation of semantic value 1 at @ in any model.

The above semantics is a generalization of the revision semantics for $\rightarrow$ of the previous section in the language $\mathcal{L}'^{+}$ without $\Rightarrow$. To see why consider a model in which W is the closed initial segment $[0, \Delta_{(v_\kappa)}]$, for some reflection ordinal $\Delta_{(v_\kappa)}$ (for example, the smallest one) for the sequence (v_κ) governing $\rightarrow$. Moreover, @ is, precisely, $\Delta_{(v_\kappa)}$ and, for each ordinal $\sigma \in W$, $\mathcal{F}_w$ has as members the intervals of the form $[\tau, \sigma)$, for all $\tau < \sigma$. An obvious problem of the model just defined is that no element of W is normal, not even @. This can be solved, though, thanks to corollary 7.2.6, for we can make @ (that is, Δ_{v_κ}) normal while leaving the rest non-normal.

We need to see, next, whether the model preserves the penumbral intuitions. As I said, each w-neighborhood is meant to represent the worlds that meet a certain standard of similarity to w. Now, suppose a and b are borderline cases (are assigned ½) of T and S at @, if we need (i) to be true at a given point w, one of the w-neighborhoods must be such that the value of Ta is always less than or equal to the value of Tb, which makes perfect sense if we understand that neighborhoods are determined by the meanings of the predicates in the language, so that at least one of the conditions of being a w-neighborhood is to be a collection of admissible and not necessarily complete

ways of making the predicates of the language precise, where admissibility is given by the meaning of the predicates and the already settled cases in w. A similar explanation can be given for claims (ii) and (iii) above so that they turn out to be true at each world in W. So the semantics can preserve the penumbral connections.

As I said, Field sees a virtue in the fact that his neighborhood semantics is a generalization of the construction for $\rightarrow$. Notice, though, that it is not clear at all that the same claim can be made with respect to the construction with $\rightarrow$ and $\Rightarrow$; recall that the value of $\rightarrow$-conditionals depends on the sequence (v_κ) we built over one of the reflection ordinals $\Omega_{(j_\kappa)}$ of the sequence (j_κ) while the value of $\Rightarrow$-conditionals depends on the sequence (j_κ). The problem is, then, that we do not know which of the series of ordinals should be assigned to the worlds in W. I do not see any natural way to remedy that.

I want to propose now another semantics for a language with a truth predicate and vague predicates that uses the models presented in section 7.2.1 and the construction used in the sections 7.2.1 and 7.2.2 for $\rightarrow$ and $\Rightarrow$. I propose to follow Field and interpret W as the possible ways to assign semantic values to the sentences in $\mathcal{L}^+$. Recall that the points in W were ordered by inclusion of the extension and the anti-extension of the predicates other than Tr, so that going up through the order can be interpreted as making some of the predicates in the language more precise; that is to say, roughly speaking, if $u \leq w$, then they both agree on the clear cases of the predicates other than Tr in u, although w may have made clear some of the cases left indetermined in u. In this framework, the conditional $\Rightarrow$, which is intended to capture the penumbral connections, works as suggested in Kamp (1981) or Shapiro (2006); a conditional is true when, no matter how you make the predicates in the conditional more precise, the consequent cannot be true without the antecedent being true. Moreover, no matter how you make the predicates in the conditional more precise, you can still precisify them in a way that the antecedent is true and the consequent is false.

Of course, this is not enough to capture claims like (i)-(iii) above, we need to restrict the points in W, that is, the possible ways of assigning semantic values to the sentences of the language. In particular, not all assignments of extensions and anti-extensions to the predicates other than Tr will be admissible, and only the admissible ones will be in W. The model, thus, is supervaluational in spirit, in the sense that uses the notion of admissible (although not necessarily complete) precisification.

Hence, when a and b are borderline cases of being 'tall' and b is taller than a, there will not be points in W where a is assigned to T^+ and b is not, so that (i) is true. Moreover, when we say that 'short' and 'tall' are contraries

what we mean is that $S^- = T^+$ and $S^+ = T^-$, that is, that the clear cases of not being short are all clear cases of being tall and that the clear cases of being short are clear cases of being not tall. If we keep these restrictions on W, (ii) and (iii) become true at any point. So that the penumbral intuition is preserved.

7.4 Solving the Paradoxes

We have seen that part of the paracomplete project consists in offering an appropriate paracomplete logic for truth. Hartry Field has been trying to do that in the last years with highly sophisticated constructions that, although succeed in strengthening SK, do not, unfortunately, give any real insight in the semantics of truth. Instead, the models are used to show that a truth predicate can be consistently (and conservatively) added to a given base model for a 'true'-free language and obtain a sufficiently strong logic. Still, we tried to see whether a model faithful to the semantics of vague terms could be given for a language with vague predicates and 'true'. We saw two proposals to do that that can be considered as constituting the same prevention both to the Liar and the *Sorites*.

On the other hand, we saw that Field defends that the linguistic practices behind 'true' and vague predicates fail to determine a unique extension and that that failure makes some of their ascriptions non-factual, which means that LEM should not be applied to them. It is when we apply LEM that we incur into a paradox.

All these considerations imply that Field is exploring a strong common solution for the Liar and the *Sorites*.[15]

[15] I need to add that I am not being completely faithful to Field's work, for he has said in many places that he does not have a "firm conviction" that a non-classical logic should be used for vagueness (see, for instance, Field 2010a, p. 458). But, still, he acknowledges that such a possibility is worth exploring, which is what I have been doing in this chapter.

CHAPTER

EIGHT

CONCLUSIONS

The main goal of this dissertation was to examine some of the major proposals for a unified account of the Liar and the *Sorites* paradoxes. In order to do that we needed to characterize, first, the notion of a common solution to a given collection of paradoxes. In turn, this needed the clarification of the notion of a solution to a single paradox, which, again, needed a definition of what a paradox is.

So, after characterizing the Liar and the *Sorites*, we looked at the traditional definition of paradox, which was found to be too narrow. We needed, hence, a more appropriate definition of what a paradox is. After discussing and discarding some alternatives, a paradox was found to be an argument that seemed valid —in the sense that rejecting its validity would imply giving up some core intuitions about the notion of logical consequence— but such that the commitment to the conclusion that stems from the acceptance of the premises and the validity of the argument should not be there. Something even stronger was stated to be the case: apparently, there is no commitment at all.

With an appropriate definition of the notion of paradox at hand we looked into what should be expected from a solution to a single paradox. We concluded that any solution should necessarily contain what Chihara called *the diagnostic* of the paradox, an explanation of the reason behind the deception in the paradox and, ideally, an explanation of why the culprit seemed so compelling in the first place. Apart from the diagnostic, a solution to a paradox might have to offer *the prevention* of the paradox, which we took to be the logico-semantic framework we need to adopt in order to block the paradox. Once we have seen all the examples of common solution present in this dissertation, it is easier now to see what a prevention of a paradox is. Note that, typically, when the need of a (non vacuous) prevention is implied by the

diagnostic, both the diagnostic and the prevention are intertwined with each other. For example, if we adopt a supervaluational approach to the *Sorites*, the prevention, which is the supervaluational semantics, is used in the diagnostic when the supervaluationist needs to explain why some premises in the *Sorites* are not true and why they are, despite that, so compelling.

Next, we were in a position to establish what should be expected from a common solution to a given collection of paradoxes. We concluded that, given a collection of paradoxes, a common solution to all of them should offer, at least, a common reason why they are paradoxical. When only such a common reason was given, we called such a solution a *weak common solution*. If, besides a common explanation of the source of paradoxicality, also a common prevention was given, then a *strong common solution* was offered. A first application of this characterization of the notion of a common solution was to see that McGee's approach to vagueness and truth in his book *Truth, Vagueness and Paradox* is not, as it stands, a common solution to the Liar and the *Sorites*. It remains to be seen whether it can be turned into a common solution, by endorsing the claim that 'true' *is* a vague predicate and not only that it *can be treated as* a vague predicate.

It was important, next, to see whether we have good reasons to expect a common solution to the Liar and the *Sorites*. We found that, apart from methodological aspects like simplicity and uniformity, one of the main reasons to pursue a common solution to some collection of paradoxes is their being of the same kind, where two paradoxes are of the same kind when they have a common reason about why they are deceptive. Although, at first sight, the Liar and the *Sorites* seem completely different paradoxes, we saw there were enough reasons to at least begin the search for a common solution. We discussed Graham Priest's criterion for being the same kind of paradox: the inclosure schema. We concluded that Priest's claim to have captured the internal structure of the *Sorites* paradox with the inclosure schema was unwarranted.

We examined, next, three proposals to cope with the Liar and the *Sorites*: Jamie Tappenden's, Paul Horwich's and Hartry Field's.

We saw that Tappenden defended the logic *SK* for languages with truth and vague predicates. In the case of truth, he accepted Kripke's fixed point construction. The problem, as we stated, was that *SK* was too weak, as it did not even validate elementary laws such as $\phi \rightarrow \phi$. The weakness of *SK* had unwelcome consequences both for truth and vague predicates; in the former case, it meant that we did not get the T-schema and, in the latter case, it meant we were not able to capture the penumbral intuitions. We looked into how Tappenden tried to handle the weakness of *SK* with supervaluationist techniques and a new speech act, which he called *articulation*. Unfortunately,

Tappenden's approach was judged to be unsuccessful, specially in the case of the Liar, where the strategy used with vague predicates turned out difficult to apply to truth. Still, I think using pragmatic considerations to deal with the inconveniences caused by embracing a non-classical logic is a clever insight worth exploring further.

Next, we looked into Horwich's stance in front of the Liar and the *Sorites*. With respect to the latter, Horwich defends an epistemicist account that accepts that vague predicates have sharp boundaries, although we cannot know them. Most of the chapter was focused on whether the epistemicist position Horwich defends for the *Sorites* can be extended to the Liar. We saw that the diagnostic offered for the Liar was in need of a prevention that used a fixed-point construction *à la* Kripke. Such a construction was made precise and was found unsatisfactory, for it did not contain some natural principles we expect any truth theory to contain. This could be solved by building into the fixed-point construction the maximal consistency of the extension of 'true'. Unfortunately, that strategy either begged the question or forced giving up deflationist views about truth. Still, it remains to explore whether a semantic version of Horwich's theory that accepted truth value gaps —and probably more supervaluationist in spirit— would be viable.

The last chapter was devoted to what I called *the paracomplete project*, which has been carried out mainly by Hartry Field. We saw a construction that used fixed-point and revision techniques in order to add two conditionals and a transparent truth predicate to a language with vague predicates. This construction, which was a variation on Field's most recent one, was a suggestion of the direction that I think the paracomplete project could follow in order to accommodate truth and vagueness. Clearly, there still is much work to be done; specially, obtaining a strong enough unified conditional. Another point I have ignored in the last chapter, and that has been deeply researched by Hartry Field, is higher order phenomena like revenge problems in the case of the Liar and higher order vagueness in the case of the *Sorites*. Field has proposed to deal with both with a operator of determinacy defined in terms of the conditional. It remains, hence, to explore this possibility in the model I proposed.

There are other projects to uniformly treat the Liar and the *Sorites* that were not treated in this dissertation. To my mind, one of the most promising ones is the treatment of both paradoxes with some substructural logic, specifically, with the restriction of the contraction rule. One such a logic is defended in Zardini (2011) to deal with the truth paradoxes and a similar strategy is proposed in Slaney (2010) in order to treat the *Sorites*. Other authors have proposed to deal with vagueness and truth with substructural logics that restrict transitivity. I think there is future work to be done in

these directions.

* * *

It is said that Philitas of Cos, a poet and philosopher of the early Hellenistic period, died of insomnia trying to solve the Liar paradox. As far as we know, nobody has ever died because of the *Sorites*. This might be due, though, to the fact that, given that there is no sharp boundary between life and death, no living being can die —if there can be living beings at all. Be that as it may, the Liar and the *Sorites* are two of the toughest paradoxes in philosophy of language, and many philosophers have struggled to solve them for centuries.

I tend to think that the we should be skeptics towards the claim that there is something like *the* solution to the Liar, or *the* solution to the *Sorites*. *A fortiori*, I think we should be skeptics towards the existence of anything like *the* common solution to both paradoxes. This would explain the failure to achieve a minimal agreement among philosophers about how they should be solved.

REFERENCES

Andjelcović, Miroslava and Timothy Williamson (2000). "Truth, Falsity, and Borderline Cases". In: *Philosophical Topics* 28.1, pp. 211–243 (cit. on p. 108).

Aquinas, Thomas (1981). *Summa Theologica*. Notre Dame: Christian Classics (cit. on p. 1).

Aristotle (1984). *The Complete Works*. Ed. by Jonathan Barnes. Princeton: Princeton University Press (cit. on pp. 1, 40, 41).

Armour-Garb, Bradley (2004). "Minimalism, the Generalization Problem and the Liar". In: *Synthese* 139.3, pp. 491–512 (cit. on pp. 107, 111).

— (2010). "Horwichian Minimalism and the Generalization Problem". In: *Analysis* 70.4, pp. 693–703 (cit. on pp. 111, 115).

Austin, J. L. (1950). "Truth". In: *Proceedings of the Aristotelian Society, supp.* 24.1, pp. 111–29 (cit. on p. 2).

Badici, Emil (2008). "The Liar Paradox and the Inclosure Schema". In: *Australasian Journal of Philosophy* 86.4, pp. 583–596 (cit. on p. 67).

Barnes, Jonathan (1982). "Medicine, Experience and Logic". In: *Science and Speculation*. Ed. by Jonathan Barnes et al. Cambridge University Press (cit. on pp. 10, 12).

Barwise, Jon and John Etchemendy (1984). *The Liar*. New York: Oxford University Press (cit. on p. 6).

Beall, J.C. (2001). "Is Yablo's Paradox Non-Circular?" In: *Analysis* 61.271, pp. 176–87 (cit. on p. 6).

Beall, J.C. and Bradley Armour-Garb (2005). "Minimalism, Epistemicism and Paradox". In: *Deflationism and Paradox*. Ed. by J. C. Beall and Bradley Armour-Garb. Oxford: Oxford University Press. Chap. 6 (cit. on p. 107).

Black, Max (1937). "Wang's Paradox". In: *Philosophy of Science* 4, pp. 427–55. Excerpts reprinted in Rosanna Keefe and Peter Smith, eds. (1997). *Vagueness: a Reader*. Cambridge, Mass.: MIT Press (cit. on p. 12).

Blackburn, Simon and Keith Simmons, eds. (2005). *Truth*. Oxford: Oxford University Press (cit. on p. 1).

Boolos, George S., John P. Burgess, and Richard C. Jeffrey (2002). *Computability and Logic*. 4th. Cambridge: Cambridge University Press (cit. on p. 8).

Booth, N. (1957). "Zeno's Paradoxes". In: *The Journal of Hellenic Studies* 77.2, pp. 187–201 (cit. on p. 41).

Bradley, F. H. (1914). *Essays on Truth and Reality*. Oxford: Clarendon Press (cit. on p. 2).

Broome, John (1999). "Normative Requirements". In: *Ratio* 12.4, pp. 398–419 (cit. on pp. 32, 33).

Burgess, Alexis G. and John P. Burgess (2010). *Truth*. Princeton: Princeton University Press (cit. on p. 1).

Cajori, Florian (1915). "The History of Zeno's arguments on Motion". In: *The American Mathematical Monthly* 22, pp. 1–6, 39–47, 77–82, 109–115, 143–149, 179–186, 215–220, 253–258, 352–357 (cit. on p. 41).

Campbell, Richmond (1974). "The Sorites Paradox". In: *Philosophical Studies* 26, pp. 175–91 (cit. on p. 18).

Cargile, James (1969). "The Sorites Paradox". In: *British Journal for the Philosophy of Science* 20, pp. 193–202 (cit. on p. 18).

Cave, Peter (2009). *This Sentence is False. An Introduction to Philosophical Paradoxes*. New York: Continuum (cit. on p. 20).

Chang, Chen Chung and H. Jerome Keisler (1973). *Model Theory*. Amsterdam: North Holland (cit. on p. 48).

Chihara, Charles S. (1979). "The Semantic Paradoxes: A Diagnostic Investigation". In: *Philosophical Review* 88.4, pp. 590–618 (cit. on pp. 44, 50).

— (1984). "The Semantic Paradoxes: Some Second Thoughts". In: *Philosophical Studies* 45.2, pp. 223–229 (cit. on p. 50).

Colyvan, Mark (2009). "Vagueness and Truth". In: *From Truth to Reality: New Essays in Logic and Metaphysics*. Ed. by H. Dyke. London: Routledge (cit. on pp. 47, 59).

Cook, Roy T. (2013). *Paradoxes*. Cambridge: Polity Press (cit. on pp. 20, 36, 40, 41, 58).

Curry, Haskell B. (1942). "The Inconsistency of certain Formal Logics". In: *Journal of Symbolic Logic* 7, pp. 115–17 (cit. on p. 6).

Dietz, Richard and Sebastiano Moruzzi, eds. (2010). *Cuts and Clouds: Vagueness, its Nature, and its Logic*. Oxford: Oxford University Press (cit. on p. 17).

Dummett, Michael (1959). "Truth". In: *Proceedings of the Aristotelian Society, n. s.* 59, pp. 141–162. Reprinted in Michael Dummett (1978). *Truth*

and other enigmas. Cambridge, Mass.: Harvard University Press (cit. on p. 2).

Dummett, Michael (1975). "Wang's Paradox". In: *Synthese* 30.3-4, pp. 201–232. Reprinted in Michael Dummett (1978). *Truth and other enigmas*. Cambridge, Mass.: Harvard University Press (cit. on pp. 13, 90).

— (1978). *Truth and other enigmas*. Cambridge, Mass.: Harvard University Press (cit. on pp. 173, 174).

Eklund, Matti (2001). "Supervaluationism, Vagueifiers, and Semantic Overdetermination". In: *Dialectica* 55.4, pp. 363–378 (cit. on pp. 92, 93).

— (2002). "Inconsistent Languages". In: *Philosophy and Phenomenological Research* 64.2, pp. 251–275 (cit. on p. 50).

Feferman, Solomon (1991). "Reflecting on Incompleteness". In: *Journal of Symbolic Logic* 56, pp. 1–49 (cit. on p. 116).

Field, Hartry (1994). "Deflationist Views of Meaning and Content". In: *Mind* 103, pp. 249–285. Reprinted in Hartry Field (2001). *Truth and the Abscence of Fact*. Oxford: Oxford University Press, pages 104–156 (cit. on pp. 2, 106).

— (2001). *Truth and the Abscence of Fact*. Oxford: Oxford University Press (cit. on pp. 116, 174).

— (2003a). "A Revenge-Immune Solution to the Semantic Paradoxes". In: *Journal of Philosophical Logic* 32.2, pp. 139–177 (cit. on pp. 145, 150, 155, 164).

— (2003b). "No Fact of the Matter". In: *Australasian Journal of Philosophy* 81.4, pp. 457–480 (cit. on pp. 138, 139).

— (2003c). "The Semantic Paradoxes and the Paradoxes of Vagueness". In: *Liars and Heaps, New Essays on Paradox*. Ed. by J.C. Beall. Oxford: Oxford University Press. Chap. 13 (cit. on pp. 18, 61, 138–140, 145, 163, 164).

— (2006). "Compositional Principles versus Schematic Reasoning". In: *The Monist* 89.1, pp. 9–27 (cit. on pp. 116, 117).

— (2007). "Solving the Paradoxes, Escaping Revenge". In: *Revenge of the Liar: New Essays on the Paradox*. Ed. by J.C. Beall. Oxford: Oxford University Press (cit. on pp. 73, 145).

— (2008). *Saving Truth From Paradox*. Oxford: Oxford University Press (cit. on pp. 18, 19, 54, 61, 68, 124–126, 134, 138, 143–145, 150, 155, 156, 163, 164).

— (2010a). "Replies to Commentators on Saving Truth from Paradox". In: *Philosophical Studies* 147.3, pp. 457–470 (cit. on p. 167).

— (2010b). "This Magic Moment: Horwich on the Boundaries of Vague Terms". In: *Cuts and Clouds: Vagueness, its Nature, and its Logic*. Ed. by

Richard Dietz and Sebastiano Moruzzi. Oxford: Oxford University Press, pp. 200 –208 (cit. on p. 110).

Field, Hartry (2014). "Naive Truth and Restricted Quantification: Saving Truth a Whole Better". In: *The Review of Symbolic Logic* 7.1, pp. 147–191 (cit. on pp. 145, 155, 156).

— (2016). "Indicative Conditionals, Restricted Quantification, and Naive Truth". In: *The Review of Symbolic Logic* 9.1, pp. 181–208 (cit. on pp. 145, 149–152, 155–159).

Fine, Kit (1975). "Vagueness, Truth and Logic". In: *Synthese* 30.3-4, pp. 265–300 (cit. on pp. 17, 48, 82).

Frege, Gottlob (1918). "Der Gedanke: Eine logische Untersuchung". In: *Beiträge zur Philosophie des deutsches Idealismus* 1, pp. 58–77 (cit. on p. 2).

Gauker, Christopher (1999). "Deflationism and Logic". In: *Facta Philosophica* 1, pp. 167–199 (cit. on p. 119).

Glanzberg, Michael (2014). "Truth". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Fall 2014 (cit. on p. 1).

Gödel, Kurt (1931). "On formally undecidable propositions of *Principia Mathematica* and related systems I". In: Gödel (1986) (cit. on p. 8).

— (1986). *Collected Works, Volume I: Publications 1929-1936*. Oxford: Oxford University Press (cit. on p. 175).

Graff Fara, Delia (2000). "Shifting Sands". In: *Philosophical Topics* 28.1, pp. 45–81 (cit. on pp. 88, 89).

Graff Fara, Delia and Timothy Williamson, eds. (2002). *Vagueness*. Aldershot: Ashgate (cit. on p. 17).

Grattan-Guinness, Ivor (1998). "Structural Similarity or Structuralism? Comments on Priest's Analysis of Paradoxes of Self-Reference". In: *Mind* 107.428, pp. 823–834 (cit. on p. 67).

Grice, H. P. (1975). "Logic and Conversation". In: *Syntax and Semantics, 3: Speech Acts*. Ed. by P. Cole and J. Morgan. New York: Academic Press (cit. on p. 87).

Grover, Dorothy (1992). *A Prosentential theory of truth*. Princeton: Princeton University Press (cit. on p. 2).

— (2001). "The Prosentential Theory: Further Reflections on Locating Our Interest in Truth". In: *The Nature of Truth*. Ed. by Michael P. Lynch. Cambridge, Mass.: MIT Press, pp. 505–526 (cit. on p. 2).

Gupta, Anil (1993a). "A Critique of Deflationism". In: *Philosophical Topics* 21.2, pp. 57–81 (cit. on pp. 105, 111).

— (1993b). "Minimalism". In: *Philosophical Perspectives* 7, pp. 359–69 (cit. on p. 111).

Gupta, Anil and Nuel Belnap (1993). *The Revision Theory of Truth*. Cambridge, Mass.: MIT Press (cit. on pp. 19, 150, 151).

Hajek, Petr, Jeff Paris, and John Sheperdson (2000). "The Liar Paradox and Fuzzy Logic". In: *Journal of Symbolic Logic* 65.1, pp. 339–346 (cit. on p. 145).

Halbach, Volker (2011). *Axiomatic Theories of Truth*. Cambridge: Cambridge University Press (cit. on p. 113).

Halldén, Sören (1949). *The Logic of Nonsense*. Uppsala: Uppsala Universitets Arsskrift (cit. on p. 18).

Herzberger, Hans G. (1970). "Paradoxes of Grounding in Semantics". In: *Journal of Philosophy* 67.6, pp. 145–167 (cit. on p. 76).

Horgan, Terence (1994). "Robust Vagueness and the Forced-March Sorites". In: *Philosophical Perspectives* 8, pp. 159–88 (cit. on p. 11).

Horsten, Leon (2011). *The Tarskian Turn*. Cambridge, Mass.: MIT Press (cit. on p. 113).

Horwich, Paul (1997). "The Nature of Vagueness". In: *Philosophy and Phenomenological Research* 57, pp. 929–36 (cit. on pp. 18, 101–103).

— (1998a). *Meaning*. Oxford: Oxford University Press (cit. on pp. 103, 109).

— (1998b). *Truth*. 2nd. Oxford: Oxford University Press (cit. on pp. 2, 105–107, 113, 117, 118, 134, 136).

— (2001). "A Defense of Minimalism". In: *Synthese* 126.1-2, pp. 149–65. Reprinted in Paul Horwich (2010c). *Truth-Meaning-Reality*. Oxford: Oxford University Press, pages 35–56 (cit. on p. 106).

— (2005a). *Reflections on Meaning*. Oxford: Oxford University Press (cit. on pp. 101, 103, 110).

— (2005b). "The Sharpness of Vague Terms". In: *Reflections on Meaning*. Oxford: Oxford University Press. Chap. 4 (cit. on p. 18).

— (2010a). "A Minimalist Critique of Tarski". In: *Truth-Meaning-Reality*. Oxford: Oxford University Press, pp. 79–97 (cit. on pp. 120, 121, 131, 135).

— (2010b). "The Notion of Paradox". In: *Truth-Meaning-Reality*. Oxford: Oxford University Press. Chap. 11 (cit. on p. 27).

— (2010c). *Truth-Meaning-Reality*. Oxford: Oxford University Press (cit. on pp. 106, 107, 114, 117, 176).

Hyde, Dominic (2011). "The Sorites Paradox". In: *Vagueness: a Guide*. Ed. by Giuseppina Ronzitti. New York: Springer (cit. on p. 12).

— (2013). "Are the Sorites and Liar Paradox of a Kind?" In: *Paraconsistency: Logic and Applications*. Ed. by Francesco Berto et al. Springer, pp. 349–366 (cit. on pp. 56, 61).

— (2014). "Sorites Paradox". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Winter 2014 (cit. on p. 12).

Jackson, Frank (1984). "Petitio and the Purpose of Arguing". In: *Pacific Philosophical Quarterly* 65, pp. 26–36. Reprinted in Frank Jackson (1987). *Conditionals*. Oxford: Basil Blackwell (cit. on p. 29).

— (1987). *Conditionals*. Oxford: Basil Blackwell (cit. on p. 177).

James, William (1907). *Pragmatism: A New Name for Some Old Ways of Thinking*. New York: Longmans, Green (cit. on p. 2).

Johnson-Laird, P.N. and Fabien Savary (1999). "Illusory inferences: a novel class of erroneous deductions". In: *Cognition* 71, pp. 191–229 (cit. on pp. 21, 22).

Kamp, Hans (1981). "The Paradox of the Heap". In: *Aspects of Philosophical Logic*. Ed. by Uwe Münich. Dordrecht: D. Reidel (cit. on p. 166).

Keefe, Rosanna (2000). *Theories of Vagueness*. Cambridge: Cambridge University Press (cit. on pp. 11, 17, 49, 87, 90, 144).

Keefe, Rosanna and Peter Smith, eds. (1997). *Vagueness: a Reader*. Cambridge, Mass.: MIT Press (cit. on pp. 17, 172).

Kirkham, Richard L. (1995). *Theories of Truth: A Critical Introduction*. Cambridge, Mass.: MIT Press (cit. on p. 1).

Körner, Stephan (1960). *The Philosophy of Mathematics*. London: Hutchinson (cit. on p. 18).

Kremer, Michael (1988). "Kripke and the Logic of Truth". In: *Journal of Philosophical Logic* 17.3, pp. 225–278 (cit. on pp. 76, 78).

Kripke, Saul A. (1975). "Outline of a Theory of Truth". In: *Journal of Philosophy* 72.19, pp. 690–716 (cit. on pp. 3, 72, 74, 76, 77, 97, 121–126, 134, 148, 149).

Künne, Wolfgang (2003). *Conceptions of Truth*. Oxford: Clarendon Press (cit. on p. 1).

Laertius, Diogenes (1972). *Lives of Eminent Philosophers*. Ed. by R. D. Hicks. Vol. 1. Cambridge, Mass.: Harvard University Press (cit. on p. 3).

Lewis, David (1974). *Counterfactuals*. Cambridge, Mass.: Harvard University Press (cit. on p. 156).

López de Sa, Dan (2009). "Can One Get Bivalence from (Tarskian) Truth and Falsity?" In: *Canadian Journal of Philosophy* 39.2, pp. 273–282 (cit. on p. 108).

López de Sa, Dan and Elia Zardini (2007). "Truthmakers, Knowledge and Paradox". In: *Analysis* 67.3, pp. 242–50 (cit. on pp. 21, 29, 31).

Lycan, W. G. (2010). "What, Exactly, is a Paradox?" In: *Analysis* 70.4, pp. 615–622 (cit. on pp. 25, 26).

Lynch, Michael P., ed. (2001). *The Nature of Truth: Classic and Contemporary Perspectives*. Cambridge, Mass.: MIT Press (cit. on p. 1).

Machina, Kenton F. (1976). "Truth, Belief and Vagueness". In: *Journal of Philosophical Logic* 5, pp. 47–78 (cit. on pp. 18, 144).

Mackie, John L. (1973). *Truth, Probability and Paradox*. Oxford: Oxford University Press (cit. on p. 1).

Makinson, David C. (1965). "The Paradox of the Preface". In: *Analysis* 25.6, pp. 205–7 (cit. on p. 37).

McGee, Vann (1989). "Applying Kripke's Theory of Truth". In: *The Journal of Philosophy* 86.10, pp. 530–539 (cit. on pp. 49, 50, 52, 54).

— (1991). *Truth, Vagueness, and Paradox*. Indianapolis: Hackett Publishing Company (cit. on pp. 47–54, 61, 124).

— (1992). "Maximal Consistent Sets of Instances of Tarski's Schema (T)". In: *Journal of Philosophical Logic* 21.3, pp. 235–241 (cit. on pp. 118, 119).

Montague, Richard (1963). "Syntactic Treatments of Modality, with Corollaries on Reflection Principles and Finite Axiomatizability". In: *Acta Philosophica Fennica* 16, pp. 153–167 (cit. on p. 9).

Patterson, Douglas (2009). "Inconsistency Theories of Semantic Paradox". In: *Philosophy and Phenomenological Research* 79.2, pp. 387–422 (cit. on p. 50).

Priest, Graham (1979). "The Logic of Paradox". In: *Journal of Philosophical Logic* 8.1, pp. 219–241 (cit. on p. 68).

— (1991). "Sorites and Identity". In: *Logique and Analyse* 34, pp. 193–96 (cit. on p. 14).

— (1994). "The Structure of the Paradoxes of Self-Reference". In: *Mind* 103.409, pp. 25–34 (cit. on pp. 57, 61, 63).

— (1997). "Yablo's Paradox". In: *Analysis* 57.4, pp. 236–242 (cit. on p. 6).

— (2000). "On the Principle of Uniform Solution: A Reply to Smith". In: *Mind* 109.433, pp. 123–126 (cit. on pp. 65, 66).

— (2002). *Beyond the Limits of Thought*. Oxford: Oxford University Press (cit. on pp. 57, 61, 63, 66).

— (2006). *In Contradiction: A Study of the Transconsistent*. Oxford: Oxford University Press (cit. on pp. 17, 21, 68, 69).

— (2008). *An Introduction to Non-Classical Logic: From If to Is*. 2nd. Cambridge: Cambridge University Press (cit. on pp. 69, 146).

— (2010a). "Inclosures, Vagueness, and Self-Reference". In: *Notre Dame Journal of Formal Logic* 51.1, pp. 69–84 (cit. on pp. 13, 61, 62, 68, 69).

— (2010b). "Non-Transitive Identity". In: *Cuts and Clouds: Vagueness, its Nature, and its Logic*. Ed. by Richard Dietz and Sebastiano Moruzzi. Oxford: Oxford University Press, pp. 200 –208 (cit. on p. 14).

Priest, Graham, J.C. Beall, and Bradley Armour-Garb, eds. (2004). *The Law on Non-Contradiction*. Oxford: Clarendon Press (cit. on p. 68).

Quine, W. V. (1966). "The ways of paradox". In: *The Way of Paradox and Other Essays*. Ed. by W. V. Quine. New York: Random House (cit. on p. 20).

Quine, W. V. (1986). *Philosophy of Logic*. 2nd. Cambridge, Mass.: Harvard University Press (cit. on pp. 104, 109).

Raatikainen, Panu (2005). "On Horwich's Way Out". In: *Analysis* 65.287, pp. 175–177 (cit. on pp. 111, 113).

Rescher, Nicholas (2001). *Paradoxes. Their Roots, Range and Resolution*. Chicago: Open Court (cit. on p. 27).

Restall, Greg (1994). *Arithmetic and Truth in Łukasiewicz's Infinitely Valued Logic*. Tech. rep. TR-ARP-6-94. Automated Reasoning Project, Australian National University (cit. on p. 145).

— (2005). "Minimalists about Truth Can (and Should) Be Epistemicists, and it Helps if They Are Revision Theorists too". In: *Deflationism and Paradox*. Ed. by J.C. Beall and Bradley Armour-Garb. Oxford: Oxford University Press. Chap. 7 (cit. on p. 107).

Ronzitti, Giuseppina, ed. (2011). *Vagueness: a Guide*. New York: Springer (cit. on p. 17).

Russell, Bertrand (1950). *The Problems of Philosophy*. Oxford: Oxford University Press (cit. on pp. 1, 2).

Sainsbury, R. M. (2009). *Paradoxes*. 2nd. Cambridge: Cambridge University Press (cit. on pp. 27, 35).

Salmon, Wesley C., ed. (2001). *Zeno's Paradoxes*. Indianapolis: Hackett Publishing Company (cit. on p. 41).

Scharp, Kevin (2013). *Replacing Truth*. Oxford: Oxford University Press (cit. on p. 50).

Schiffer, Stephen R. (2003). *The Things we mean*. Oxford: Clarendon Press (cit. on pp. 27, 42, 43).

Shapiro, Stewart (2006). *Vagueness in Context*. Oxford: Oxford University Press (cit. on pp. 164, 166).

Simmons, Keith (1993). *Universality and the Liar. An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press (cit. on p. 52).

Sinnott-Armstrong, Walter (2012). "Begging the Question". In: *Australasian Journal of Philosophy* 77.2, pp. 174–91 (cit. on p. 30).

Slaney, John K. (2010). "A Logic fro Vagueness". In: *Australasian Journal of Logic* 8, pp. 100–134 (cit. on p. 170).

Smith, Nicholas J. J. (2000). "The Principle of Uniform Solution (of the Paradoxes of Self-Reference)". In: *Mind* 109.433, pp. 117–122 (cit. on p. 64).

— (2008). *Vagueness and Degress of Truth*. Oxford: Oxford University Press (cit. on pp. 18, 144).

Smith, Peter (2007). *An Introduction to Gödel's Theorems*. First edition. Cambridge: Cambridge University Press (cit. on p. 8).

Soames, Scott (1997). "The Truth About Deflationism". In: *Philosophical Issues,* 8*, Truth*. Ed. by Enrique Villanueva. Atascadero, California: Ridgeview Publishing Company, pp. 1–44 (cit. on p. 111).
— (1999). *Understanding Truth*. Oxford: Oxford University Press (cit. on pp. 21, 78, 111, 112).
Sorensen, Roy A. (1988). *Blindspots*. Oxford: Clarendon Press (cit. on p. 18).
— (1998). "Yablo's Paradox and Kindred Infinite Liars". In: *Mind* 107.425, pp. 137–155 (cit. on p. 6).
— (2001). *Vagueness and Contradiction*. Oxford: Oxford University Press (cit. on pp. 17, 18).
— (2016). "Vagueness". In: *The Stanford Encyclopedia of Philosophy*. Ed. by Edward N. Zalta. Spring 2016 (cit. on p. 17).
Stalnaker, Robert C. (1968). "A Theory of Conditionals". In: *Americal Philosophical Quarterly*, pp. 98–112 (cit. on p. 156).
Tappenden, Jamie (1993). "The Liar and Sorites Paradoxes: Toward a Unified Treatment". In: *Journal of Philosophy* 60.11, pp. 551–577 (cit. on pp. 61, 72, 81, 82, 85, 86, 88, 91, 96, 97).
Tarski, Alfred (1944). "The Semantic Conception of Truth: And the Foundations of Semantics". In: *Philosophy and Phenomenological Research* 4.3, pp. 341–376 (cit. on pp. 7, 51).
— (1969). "Truth and Proof". In: *Scientific American* , June, pp. 63–70, 75–77 (cit. on pp. 44, 51).
— (1983). "The Concept of Truth in Formalized Languages". In: *Logic, Semantics, Metamathematics*. Ed. by John Corcoran. Trans. by J. H. Woodger. 2nd. Indianapolis: Hackett Publishing Company, pp. 152–278 (cit. on pp. 7, 50, 51, 111).
Tye, Michael (1994). "Sorites Paradoxes and the Semantics of Vagueness". In: *Philosophical Perspectives* 8, pp. 189–206 (cit. on p. 18).
Valor Abad, Jordi (2008). "The Inclosure Scheme and the Solution to the Paradoxes of Self-Reference". In: *Synthese* 160, pp. 183–202 (cit. on pp. 66, 67).
van Fraassen, Bas C. (1966). "Singular Terms, Truth Value Gaps and Free Logic". In: *Journal of Philosophy* 63, pp. 481–495 (cit. on pp. 17, 48).
Visser, Albert (1989). "Semantics and the Liar Paradox". In: *Handbook of Philosophical Logic* 4.1, pp. 617–706 (cit. on p. 25).
Vlastos, Gregory (1967). "Zeno of Elea". In: *The Encyclopedia of Philosophy*. Ed. by Paul Edwards. New York: The Macmillan Company and The Free Press (cit. on p. 41).
Weber, Zach and Mark Colyvan (2010). "A Topological Sorites". In: *The Journal of Philosophy* 107, pp. 311–925 (cit. on p. 16).

Weir, Alan (1996). "Ultramaximalist Minimalism!" In: *Analysis* 56.1, pp. 10–8211 (cit. on p. 119).

Williamson, Timothy (1994). *Vagueness*. London: Routledge (cit. on pp. 17, 18, 49).

Wright, Crispin (1975). "On the Coherence of Vague Predicates". In: *Synthese* 30, pp. 325–65 (cit. on p. 11).

Yablo, Stephen (1989). "Truth, Definite Truth and Paradox". In: *The Journal of Philosophy* 86.10, pp. 539–541 (cit. on p. 52).

— (1993). "Paradox Without Self–Reference". In: *Analysis* 53.4, pp. 251–252 (cit. on pp. 5, 6).

— (2004). "Circularity and Paradox". In: *Self-Reference*. Ed. by Thomas Bolander, Vincent F. Hendricks, and Stig Andur Pedersen. Csli Publications, pp. 139–157 (cit. on p. 6).

Young, James O. (1995). *Global Anti-realism*. Aldershot: Avebury (cit. on p. 2).

Zadeh, Lofti (1975). "Fuzzy Logic and Approximate Reasoning". In: *Synthese* 30, pp. 407–428 (cit. on p. 18).

Zardini, Elia (2011). "Truth Without Contra(di)ction". In: *Review of Symbolic Logic* 4.4, pp. 498–535 (cit. on pp. 19, 170).

Zhong, Haixia (2012). "Definability and the Structure of Logical Paradoxes". In: *Australasian Journal of Philosophy* 90.4, pp. 779–788 (cit. on p. 67).