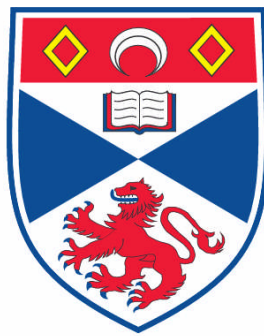


**LIVING ON THE SLIPPERY SLOPE : THE NATURE, SOURCES
AND LOGIC OF VAGUENESS**

Elia Zardini

**A Thesis Submitted for the Degree of PhD
at the
University of St. Andrews**



2008

**Full metadata for this item is available in the St Andrews
Digital Research Repository
at:**

<https://research-repository.st-andrews.ac.uk/>

Please use this identifier to cite or link to this item:

<http://hdl.handle.net/10023/508>

This item is protected by original copyright

**This item is licensed under a
[Creative Commons License](#)**

Living on the Slippery Slope

The Nature, Sources and Logic of Vagueness

Elia Zardini

Arché, Philosophical Research Centre for Logic, Language, Metaphysics and Epistemology

Department of Logic and Metaphysics

University of St Andrews

`ez4@st-andrews.ac.uk`

Thesis submitted for the degree of PhD in Philosophy

November 22, 2007

Abstract

According to the dominant approach in the theory of vagueness, the nature of the vagueness of an expression ‘ F ’ consists in its presenting borderline cases in an appropriately ordered series: objects which are neither definitely F nor definitely not F (where the notion of definiteness can be semantic, ontic, epistemic, psychological or primitive). In view of the various problems faced by theories of vagueness adopting the dominant approach, the thesis proposes to reconsider the naive theory of vagueness, according to which the nature of the vagueness of an expression consists in its not drawing boundaries between any neighbouring objects in an appropriately ordered series. It is argued that expressions and concepts which do present this feature play an essential role in our cognitive and practical life, allowing us to conceptualize—in a way which would otherwise be impossible—the typically coarse-grained distinctions we encounter in reality. Despite its strong initial plausibility and ability to explain many phenomena of vagueness, the naive theory is widely rejected because thought to be shown inconsistent by the sorites paradox. In reply, it is first argued that accounts of vagueness based on the dominant approach are themselves subject to higher-order sorites paradoxes. The paradox is then solved on behalf of the naive theory by rejecting the unrestricted transitivity of the consequence relation on a vague language; a family of logics apt for reasoning with vague expressions is proposed and studied (using models with partially ordered values). The characteristic philosophical and logical consequences of this novel solution are developed and defended in detail. In particular, it is shown how the analysis of what happens in the attempt of surveying a sorites series and deciding each case allows the naive theory to recover a “thin” notion of a borderline case.

Declarations

I, Elia Zardini, hereby certify that this thesis, which is approximately 90,000 words in length, has been written by me, that it is the record of work carried out by me and that it has not been submitted in any previous application for a higher degree.

Date: 22/11/2007

Signature of candidate

I was admitted as a research student in September, 2003 and as a candidate for the degree of PhD in September, 2003; the higher study for which this is a record was carried out in the University of St Andrews between 2003 and 2007.

Date: 22/11/2007

Signature of candidate

I hereby certify that the candidate has fulfilled the conditions of the Resolution and Regulations appropriate for the degree of PhD in the University of St Andrews and that the candidate is qualified to submit this thesis in application for that degree.

Date: 22/11/2007

Signature of supervisor

In submitting this thesis to the University of St Andrews I understand that I am giving permission for it to be made available for use in accordance with the regulations of the University Library for the time being in force, subject to any copyright vested in the work not being affected thereby. I also understand that the title and abstract will be published, and that a copy of the work may be made and supplied to any bona fide library or research worker, that my thesis will be electronically accessible for personal or research use, and that the library has the right to migrate my thesis into

new electronic forms as required to ensure continued access to the thesis. I have obtained any third-party copyright permissions that may be required in order to allow such access and migration.

Date: 22/11/2007

Signature of candidate

v

Per Zeno

Acknowledgements

Without dialogue, philosophy would be no more than a raving dream. Many people have contributed to shaping the ideas presented in this essay. I first want to register special debts of gratitude to five persons, thanks to whom I can stutter some philosophy. Lucio Regaiolo was my first teacher in philosophy at high school in my hometown, from 1995 to 1998. I had actually already decided to become a philosopher before that, but, at that age, anything less than a terrific and inspiring teacher would have easily caused me to deflect from that crazy intention. I didn't deflect. Lucio passed away early this year, to my deep regret for the conversations that I still would have liked to have with him but that I didn't have. Emanuele Severino taught me philosophy during my undergraduate study in Venice from 1998 to 2003. Through his tantalizing lectures and writings, I first understood what it really is to do philosophy and what philosophy can be at its best. Sven Rosenkranz first introduced me—among many other issues—to the problem of vagueness during my Erasmus visit in Berlin from 2001 to 2002. Without Sven's crucial guidance on those first steps, this essay—and many other things—would simply not exist. I'm indebted to him for this—and for much more. Crispin Wright has been patient enough to be my supervisor for four years. Anyone acquainted with Crispin knows what it is like to have a philosophical conversation with him—an intellectual excitement whose beneficial effects linger on for the whole day. And anyone acquainted with his ideas can tell how deeply they have influenced these pages, and not just them. He has believed in me and in this project: I can hardly imagine what sort of philosopher I would have ended up being without his help and advice. Over the years, Crispin has also been a constant example of intellectual depth, rigour, passion and honesty. I am exceedingly proud to have him as "*Doktorvater*". Patrick Greenough has also been patient enough to be my supervisor for four years. He's the only person in the world capable of giving a five and a half hour tutorial and one of the few ready to continue a philosophical discussion with

an insatiable supervisee after a hearty Czech dinner. His open-mindedness, receptiveness and support have been terrific and I have learnt from him much of what I know about vagueness and many other philosophical topics.

I have conducted most of this research in Arché, a research centre for “hard-core” analytic philosophy. It’s a place where one can go on the same day from talking about the foundations of mathematics to talking about scepticism, with a little digression on the meaning of ‘I’. A philosopher’s paradise. I have immensely benefitted from exhausting discussions on vagueness with just about every (formal or informal) associate of the centre: Jonas Åkerman, JC Beall, Eline Busck, Ross Cameron, Herman Cappelen, Roy Cook, Neil Cooper, Philip Ebert, Douglas Edwards, Kit Fine, Bob Hale, Katherine Hawley, Richard Heck, Andri Hjálmarsson, Ole Hjortland, Aviv Hoffmann, Carrie Jenkins, Ira Kiourti, Michael Lynch, Roberto Loss, Federico Luzzi, Darren McDonald, Aidan McGlynn, Paul McCallion, Paula Milne, Daniel Nolan, Josh Parsons, Carlotta Pavese, Nikolaj Pedersen, Walter Pedriali, Agustín Rayo, Stephen Read, Sònia Roca, Marcus Rossberg, Daniele Sgaravatti, Stewart Shapiro, Martin Smith, Andreas Stokke, Chiara Tabet, Jordi Valor, Brian Weatherson and Robbie Williams. Special thanks go to the members of the Arché Vagueness Project, which ran from 2004 to 2006 and has been an incredible intellectual adventure. These include—in addition to some people already mentioned—Elizabeth Barnes, Richard Dietz, Dan López de Sa and Sebastiano Moruzzi. Mark Sainsbury has been auditing the project for four years, always giving very careful, detailed and constructive feed-back on my work-in-progress. Graham Priest has supervised me in the second half of 2006: I have learnt much from Graham’s extraordinary broad-mindedness and sharp experimental spirit, which have been a great inspiration in developing my own logical heterodoxy. Thanks are due also to the various secretaries who have made College Street 17–9 a wonderful working environment and have helped with lots of enigmatic *administrativa*: Gill Gardner, Silvia Rescigno and especially Sharon Coull, who among other uniquely identifying properties has also that of being the best Office Manager

in the world.

I was visiting scholar at the NYU philosophy department in the first half of 2006. I'm very grateful to Crispin Wright and Stephen Schiffer for arranging the visit, which provided me with a tremendous intellectual input. Stephen's critical sharpness has led me over the years to reconsider and modify important aspects of my views on vagueness, and I've always admired the striving for truth manifested in his proverbially intensive questioning. During my stay at NYU, I was lucky enough to have as supervisor someone who's got exactly the same fixation with heaps and liars: I'm extremely thankful to Hartry Field for very illuminating conversations on these topics. The systematicity and thoroughness of Hartry's approach to these most difficult issues, together with its unique blend of philosophical insight and technical prowess, have been a constant model for my own thinking.

Material from the second chapter has been presented in 2006 at the Arché Vagueness Seminar and at the Third Navarre Vagueness Workshop (University of Navarre, in Granada). Material from the third chapter has been presented in 2004 at the Arché Vagueness Seminar; in 2005 at the 9th NPAPC (University of York), where David Eford gave a valuable response; in 2006 at a Vagueness Workshop (University of Freiburg, in Oronnaz), where Reto Givel gave another valuable response, and at the 6th GAP Conference on *Philosophy: Foundations and Applications* (Berlin Free University); in 2007 at the 1st UCLA/USC Graduate Student Conference in Philosophy (UCLA), where Gabe Rabin gave yet another valuable response. Material from the fourth chapter has been presented in 2006 at the Prague International Colloquium on *Uncertainty: Reasoning about Probability and Vagueness* (University of Prague); in 2007, at the Arché Vagueness Seminar, at 13th International Congress of Logic, Methodology and Philosophy of Science (Qīnghuá University, Beijing), at the 2nd World Congress on Universal Logic (Northwestern Polytechnical University, Xī'ān) and at the SEFA Conference (University of Barcelona).

I'm grateful to all these and other audiences and philosophers around the

world for extremely helpful feed-back on my ideas on vagueness. Apologizing in advance for the omissions which here are going to be most likely, I'd like to mention Libor Běhounek, Jiri Benovsky, Hanoch Ben-Yami, Jean-Yves Béziau, Ross Brady, María Cerezo, Pablo Cobreros, Annalisa Coliva, Josh Dever, Cian Dorr, Kenny Easwaran, Dorothy Edgington, Paul Égré, Matti Eklund, Delia Fara, Manuel García-Carpintero, Mikkel Gerken, Laurence Goldstein, John Hawthorne, Mark Heller, Terence Horgan, Dominic Hyde, Andrea Iacona, Rosanna Keefe, Philipp Keller, Max Kölbel, Jonathan Lowe, John MacFarlane, Teresa Marques, Andrew McGonigal, Bob Meyer, Julien Murzi, Peter Pagin, Marco Panza, Philip Percival, Diana Raffman, Greg Restall, Enrique Romerales, Peter Simons, Roy Sorensen, Giuliano Torrenco, Achille Varzi and Tim Williamson.

Darren McDonald offered much needed editing assistance with the diagrams of chapter 4. Patrick Greenough and Stephen Read read a penultimate draft of this essay, providing me with excellent feed-back on both substance and form. Large parts of a final draft have been painstakingly read through by my friends Laura Delgado, Ira Kiourti, Walter Pedriali, Andri Hjálmarsson, Marcus Rossberg, Kevin Scharp and Douglas Edwards. I'm extremely grateful to all of them for their generous help.

In writing this essay, I have benefitted from an Arts and Humanities Research Council Doctoral Award from 2004 to 2006 and from a Royal Institute of Philosophy Jacobsen Fellowship from 2006 to 2007. I'm very grateful to both institutions for their financial support.

I have left out “acknowledgements” regarding my extra-academic life in the last four years—there would be many, and most persons involved have not been mentioned here. But I wish to dedicate this essay to Zeno, good friend, who has taught me much about all that. *Tat tvam asi.*

Contents

1	Introduction	1
1.1	Introduction and Overview	1
1.2	The Bearers of Vagueness	2
1.3	The Phenomena of Vagueness	3
1.4	The Theories of Vagueness	11
1.5	Plan	18
2	Seconde Naïveté	21
2.1	Introduction and Overview	21
2.2	Thoughts Requiring the Absence of Sharp Boundaries	30
2.2.1	Irrelevant Differences	30
2.2.2	Stretching the Truth	39
2.2.3	Application by Casual Observation	51
2.3	Experiences Requiring the Absence of Sharp Boundaries	57
2.3.1	Seamless Change	57
2.3.2	Appearances	65
2.4	Conclusion	71
3	Higher-Order Sorites Paradox	73

3.1	Introduction and Overview	73
3.2	Ignorance	74
3.2.1	Ignorance and Borderlineness	74
3.2.2	Indefiniteness and Tolerance	76
3.3	Higher-Order Vagueness	78
3.3.1	Being Definite ^ω	78
3.3.2	Radical Higher-Order Vagueness	83
3.4	The Paradox	84
3.4.1	The Basic Form of the Paradox	84
3.4.2	Weakening the Logic?	87
3.5	Conclusion	90
4	Tolerant Logics	93
4.1	Introduction and Overview	93
4.2	A Neutral Framework	97
4.2.1	Syntax	97
4.2.2	Tolerant Semantic Structures	99
4.2.3	Basic Tolerant Logic	107
4.3	Classical Tolerant Logic	119
4.3.1	Towards the Classical Tolerant Logic CT ⁰	119
4.3.2	The Strength of CT ⁰	121
4.3.3	The Weakness of CT ⁰	127
4.4	The Consistency of the Naive Theory of Vagueness	135
4.4.1	Consistency in a Tolerant Framework	135
4.4.2	A Model of Tolerance	138
4.5	Going First-Order	139

4.5.1	Vagueness and First-Order Expressive Power	139
4.5.2	Syntax	142
4.5.3	Tolerant First-Order Semantic Structures	145
4.5.4	First-Order Basic Tolerant Logic	147
4.5.5	First-Order Classical Tolerant Logic	150
4.5.6	Naive Abstraction	154
4.5.7	The Consistency and Strength of the First-Order Naive Theory of Vagueness	157
4.5.8	The Price of Tolerance	164
4.6	Conclusion	173
5	Following-from and Transitivity	175
5.1	Introduction and Overview	175
5.2	Non-Transitive Consequence Relations	180
5.2.1	Relevance	180
5.2.2	Tolerance	185
5.2.3	Probabilistic Reasoning	186
5.3	Two Objections to the Very Idea of Non-Transitive Consequence	189
5.3.1	Consequence and Inference	189
5.3.2	Consequence and Truth Preservation	193
5.4	The Non-Transitivist's Picture	201
5.4.1	Non-Logical/Logical Dualism and Non-Transitivism . .	201
5.4.2	Logical Nihilism and Non-Transitivism	206
5.4.3	The Normativity of Consequence	211
5.4.4	The Asymmetry between Premises and Conclusions . .	221
5.4.5	The Locality of Non-Transitive Consequence	223

5.4.6	Non-Transitivist Theories, Situations and Worlds . . .	225
5.5	Conclusion	231
6	Forced March in the Penumbra	233
6.1	Introduction and Overview	233
6.2	Preliminaries	236
6.2.1	A Simple-Minded Conjecture and Innocent Questions .	236
6.2.2	Saying ‘No’	238
6.3	Can’t Say ‘Can’t Say’	240
6.3.1	Failure of Negative Transparency	240
6.3.2	Epistemic Paradox	241
6.4	The Forced March	247
6.4.1	Setting	247
6.4.2	Sequentially Inconsistent Judgements	250
6.4.3	Sequentially Consistent Judgements	258
6.4.4	Unwise Chrysippus	264
6.4.5	The Enforcement of Classical Logic: Transparency and Inexact Knowledge	274
6.4.6	Naive Forced March	281
6.5	The Penumbra	289
6.5.1	Falling Asleep with an Unhappy Face	289
6.5.2	Doxastic Paradox	293
6.5.3	The Source of Borderline Cases	297
6.6	Conclusion	304
7	Conclusion	307

Chapter 1

Introduction

1.1 Introduction and Overview

Words like ‘bald’, ‘heap’ and ‘red’ are said to be *vague*, in opposition to words like ‘prime natural number’, ‘self-identical’ and ‘triangular’, which are said to be *precise* (i.e. non-vague). Vague words are said to be so because, very roughly, their area of application is somehow blurred, not sharply delimited—as Frege famously observed, “to represent the extensions of concepts as areas on a plane is merely an analogy which should be employed with caution, but which can be useful [...] [a]n unsharply delimited concept would be associated with an area which, instead of having a sharp boundary everywhere, at some places blurs into its surroundings” (Frege [1903], §56, my translation; see Dummett [1981], pp. 33–5; van Heijenoort [1986]; Williamson [1994], pp. 37–46 for discussions of Frege’s views on vagueness). My main questions will be: What does this blur consist in? Why does it arise? How should we reason in its presence?

This essay offers a new theory of vagueness. The theory needs some stage setting before being properly introduced and this introduction is organized accordingly. Section 2 delineates what the bearers of vagueness are with which this essay will be concerned. Section 3 describes the main phenomena

of vagueness. Section 4 sketches two different approaches to theorizing about these phenomena and introduces the theory to be developed and defended in the rest of the essay. Section 4 outlines what this rest will amount to.

1.2 The Bearers of Vagueness

Vagueness is understood primarily to be a *semantic* property of *linguistic expressions* (that is, a property concerning certain aspects of their *meaning* rather than of their *form*), like ambiguity, abstractness, rigidity etc. (if the language in question is context dependent, it might however be better to attach vagueness to *particular events of use* of a certain linguistic expression). As in many other cases of philosophical interest, *language* finds its counterpart in *thought*, so that it is very plausible to assume that some concepts exhibit features analogous to those exhibited by vague words and, again as in many other cases of philosophical interest, light on this feature of human thought can be hoped to be shed by investigation of the public languages which serve to express it. I will generally follow this methodological principle, focussing mainly on the vagueness of words, but sometimes shifting to the vagueness of concepts where it seems appropriate—my main claims and arguments will be general enough as to apply to both language and thought, and hopefully also to other systems of representation, like for example pictures.

As for language, I should also stress from the start that I will make the gross oversimplification of pretending that we speak a standard first-order language, treating English common nouns, adjectives, verbs etc. as simple predicates of such language. I will also consider only such predicates as possible bearers of vagueness, leaving for another occasion the extension of the theory to linguistic expressions belonging to other semantic categories like singular terms, quantifiers, operators etc. (even though the logical framework developed in chapter 4 already contains important elements for the treatment

of the vagueness of singular terms). With so much by way of preliminaries about the bearers of vagueness, we can turn to vagueness itself (see Rolf [1981], pp. 150–1; Burns [1991], pp. 7–16 for more discussion on the bearers of vagueness).

1.3 The Phenomena of Vagueness

There has been a considerable debate in the literature as to how broadly the word ‘vagueness’ is and should be used in theorizing about language and, relatedly, as to what the relations are between “vagueness” and other traditional semantic properties of linguistic expressions such as ambiguity, underspecificity, context dependence etc. (see Alston [1964], p. 85; Sorensen [1989]; Sorensen [1998]; Rolf [1981], pp. 74–7; Burns [1991], pp. 16–20; Burns [1995]; Keefe and Smith [1997], pp. 5–6; Keefe [2000], pp. 10–1). This essay will follow the established philosophical tradition of reserving the word ‘vagueness’ and its like for a very specific property whose phenomena we now proceed to characterize (see Greenough [2003] for a very useful in-depth discussion of these). A predicate will count as vague iff it exhibits all these phenomena. Such a characterization is supposed to be relatively uncontroversial—the task of a theory of vagueness will then be to explain what it is about a vague predicate which causes these phenomena.

Sorites Susceptibility. The first and foremost phenomenon of the vagueness of a predicate is its seeming failure to draw a sharp boundary between its positive and negative cases. For example, it seems preposterous to think that there is a sharp boundary between the numbers which are small and those which are not—that is, that there is a number i such that i is small and $i + 1$ is not small (throughout, let us understand ‘small’ in such a way that 0 is indisputably small and 1,000,000 one of the first indisputably non-small numbers).

Note that many commentators nowadays use the phrase ‘sharp boundary’

and similar ones in a different way, such that i is a “sharp boundary” between the small numbers and the non-small ones iff i is *definitely* small and $i + 1$ is *definitely* not small (see the next part on indefiniteness for more on ‘definitely’). Since the theory I will advocate in this essay has it that in this case appearances correspond to reality and hence that ‘small’ does not in effect draw a *unit-sized* boundary between its positive and negative cases, I should wish to deviate from such contemporary usage. For the qualification ‘sharp’ and similar ones convey very well the idea that the boundary which my theory denies to exist is unit-sized, whereas stating the theory using the bare phrase ‘boundary’ would too easily induce the false impression that it denies that there is any distinction at all between the small numbers and the non-small ones. To be clear then, the first and foremost phenomenon of the vagueness of ‘small’ is not that it is a *fact* that there is no number i such that i is definitely small and $i + 1$ is definitely not small, but that it *seems* that there is no number i such that i is small and $i + 1$ is not small.

What thus seems to be the case is something which generates that particular kind of slippery-slope arguments known as ‘*sorites paradoxes*’ (see chapter 4 for a presentation and ample discussion of these), which give our phenomenon its name and for which vagueness is usually identified as the culprit. In my view, this is exactly as it should be, since the premises of even the best slippery-slope arguments making use only of precise predicates like ‘ ξ bricks would break a camel’s back’¹ fall well short of the extreme plau-

¹When I started writing this essay, I was still aiming at employing an absolutely rigorous system of quotation. In the end, I have had to acknowledge that the complexity and clumsiness of the notation needed to achieve the purpose would greatly override any gain in conceptual clarity it might have served. I have thus decided to settle for a rather promiscuous system where single quotes and display both serve in context to indicate the appropriate quotation environment (simple quotation, quasi-quotation, autonymous quotation etc. and combinations thereof) and where the decision as to how to interpret the use of letters like ‘ x ’, ‘ F ’, ‘ P ’ etc. (as non-substitutional variables, or as substitutional variables, or as schematic letters etc.) is also left to context (whereas, to avoid excessive confusion, I stuck to maintaining a distinct notation for metalinguistic variables, using ‘ ξ ’, ‘ Φ ’, ‘ φ ’ etc.).

sibility and peculiar immunity to demonstrable counterexamples enjoyed by the premises of a sorites paradox.

Note also that I will assume throughout that its not being the case that x is F and y is not F is equivalent with its being the case that, if x is F , so is y . However, if one wishes to recognize relevantist distinctions, in order to deal with certain *recherché* cases (see section 1.4) ‘and’ should be understood in such contexts to denote *extensional conjunction* rather than *fusion* (Read [1988], pp. 36–50 is probably the best philosophical discussion of this distinction), so that the equivalence would break down (left-to-right).

Indefiniteness. The second phenomenon of the vagueness of a predicate has just emerged and is its failure to draw a definite sharp boundary between its positive and negative cases. For example, there is no definite sharp boundary between the numbers which are small and those which are not—that is, there is no number i such that i is definitely small and $i + 1$ is definitely not small.

Although clearly appealing to some pre-theoretical intuition, ‘definitely’ is here something of a term of art. It is best introduced ostensively by saying that 0, 1, 2 and a few other close numbers are definitely small, that 1,000,000, 1,000,001, 1,000,002 and a few other close numbers are definitely not small and that some numbers in the middle of the series, like 499,999, 500,000, 500,001 and a few other close numbers, are neither definitely small nor definitely not small. It is then a major task of a theory of vagueness using the notion of being definite to provide a theoretical explication of what the property of being definite consists in. The contemporary literature on vagueness does in effect provide a wide range of proposals both for an *explicit* definition of ‘Definitely, φ ’:

Semantic: Speakers’ practices determine a sufficient truth condition for ‘ φ ’, and the condition is satisfied (see e.g. Fine [1975]; McGee and McLaughlin [1995]);

Ontic: It is a fact of the matter that φ (see e.g. Tye [1990]);

Epistemic: No obstacle of a certain kind prevents knowledge that φ (see e.g. Sorensen [1988], pp. 199–252; Williamson [1994]; Sorensen [2001]);

Psychological: It is correct to hold a standard (possibly partial) belief to the effect that φ (see e.g. Schiffer [2003], pp. 178–237)

and for an *implicit* definition of it by means of a specification of the conceptual role of the ‘definitely’-operator (see e.g. Field [1994], pp. 409–22; Field [2003b]).²

Note that a predicate can arguably be indefinite without thereby being sorites susceptible. For example, a partial stipulation to the only effect that ‘small*’ is to apply positively from 0 to 499,998 and negatively from 500,001 onwards has arguably the result that the numbers from 0 to 499,998 are definitely small*, the numbers from 500,001 onwards definitely not small* and 499,999, 500,000 and 500,001 neither definitely small* nor definitely not small* (see Sainsbury [1991], p. 173). Such a pattern suffices to ensure indefiniteness, but not sorites susceptibility, since it does not seem to be the case that, if i is small*, so is $i + 1$ (in particular, it does not seem to be the case that, if 499,998 is small*, so is 499,999).

Borderlineness. The third phenomenon of the vagueness of a predicate is connected with indefiniteness and is its possible presentation of borderline cases.³ For example, on each episode of consideration, some numbers

²Anticipating a little, I should note that, given the usual definition of a borderline case as something which is neither a definitely positive case nor a definitely negative case, all such proposals will imply that borderline cases are strong borderline cases in the sense explained in the next part on borderlineness. Hence, for reasons I will very briefly touch on in that part, all such proposals are unacceptable on my view. My own deflationary story about borderline and definite cases is told in section 6.5.3.

³‘Borderline’ and its like can hardly be said to express a unique notion in ordinary English. For example, on some uses, a 19-year-old man is a “borderline” teenager. I’m thus going to pick up only on a fairly specific practice with the word ‘borderline’, namely the one which associates it in a peculiar way with the other phenomena of vagueness. Prominent in such a practice is the phenomenon to be presently introduced in the text,

are borderline small, in the sense that not every positive or negative ‘small’-judgement about numbers from 0 to 1,000,000 can be warranted on that episode (this being understood as leaving it open whether each of these judgements can be warranted on some episode of consideration or other). This impossibility is reflected in the fact that we would prefer not to give a positive or negative ‘small’-judgement for every number from 0 to 1,000,000 and would rather leave some cases in the middle of the series unexamined. If forced to give such a judgement for every number in the series, we only do so with a sense of absurdity and artificiality, as though being forced to trespass the limits of what is warranted (see chapter 6).

This is a minimal description of borderlineness. Many commentators believe that the phenomenon allows for stronger descriptions, manifesting, on each particular episode of consideration, a certain *distinctive normative status* (either semantic, or ontic, or epistemic, or psychological, or primitive, along the lines discussed earlier) enjoyed by some numbers in the middle of the series quite independently of the fact that, on that episode, these numbers cannot be judged on pain of the sense of absurdity and artificiality just noted. Let us refer to the possession of this alleged status as ‘*strong borderlineness*’. The very same notion of being definite, once provided with a suitable theoretical explication, is then used to describe such a status, claiming that some numbers in the middle of the series are neither definitely small nor definitely not small (indeed, the property of being (strong) borderline is usually assumed to be so defined in terms of the property of being definite and I will normally follow suit in this respect, save for section 3.3.1).⁴

but arguably the practice envisages also other features not reducible to that. I rather tentatively assume that the practice is articulated and coherent enough as to warrant the use of the definite description ‘the pre-theoretical notion of a borderline case’ (see section 6.5.3), but not much will hinge on this assumption, so that, as far as borderlineness goes, we can focus almost exclusively on the phenomenon about to be described.

⁴Greenough [2003], pp. 265–72; Cook [2005]; Zardini [2006b] contain extended analyses of the logical relations between different versions of indefiniteness and borderlineness under the usual definition of the property of being borderline. In this connection, I should stress

For reasons which I cannot hope to expand on in the limited space of this essay, I don't think that such stronger descriptions are phenomenologically warranted.⁵ To put it very briskly, it seems that nothing may be wrong with someone who, after some reflection and in acknowledgement of the epistemic difficulty of the situation, resolutely believes that 499,999 is small. Indeed, it seems that nothing may be wrong with her claiming that she knows that 499,999 is small. This simple piece of data is surprisingly inconsistent with theories which attribute strong borderlineness to 499,999 with respect to smallness. Therefore, they should be rejected.

Ignorance. The fourth phenomenon of the vagueness of a predicate escalates the epistemic features of borderlineness and is our present ignorance of a sharp boundary between its positive and negative cases (chapter 6, fn 15 critically discusses attempts at questioning this phenomenon). For example, *we* are now not able to provide a warranted identification of a sharp boundary between the small numbers and the non-small numbers *on the basis of the generally available methods* used to track smallness and non-smallness. Almost any present attempt of ours at such an identification would be groundless—and if not rejected for this reason, any alleged such

that, on my view, while the entailment from borderlineness to indefiniteness is rather unproblematic (but see section 3.4.2 for some possible qualms), the converse entailment from indefiniteness to borderlineness does not hold, since it can only be established by making use of arguments which are objectionable on the logic for vagueness that I will develop in chapter 4.

⁵I owe to some of the recent writings of Crispin Wright on this issue (Wright [2001]; Wright [2003a]; Wright [2003b]; Wright [2007a]) the general idea that the traditional philosophical picture of borderline cases as enjoying strong borderlineness is dramatically inflated and out of touch with the data. Conversations with him over the years have been invaluable in shaping my own views on the matter. Nevertheless, I should stress that the radical view expressed in the text is something that Wright has never committed himself to—indeed, while he has done much to debunk the idea that borderline cases enjoy some kind of distinctive *semantic* status, in some of his writings he shows sympathy for the idea that they still enjoy some kind of distinctive *epistemic* and *psychological* status (see especially Wright [2001]; Wright [2007a]).

identification would rather be taken as a demonstration that the predicate was not, after all, vague. Note that this is not to rule out that someone *else* is able to use these very *same* methods so as to provide such an identification nor to rule out that there are some *other* methods which would enable *us* to provide such an identification (see section 3.2.1 for further discussion).

The ignorance in question is thus understood to extend from lack of *knowledge* to a more general lack of *warrant*, although this is not to deny that, in an appropriately weak sense of ‘justification’ (sense in which ‘justification’ is roughly equivalent with ‘epistemic support of some sort or other’), one may well have a positive degree of justification in the identification of a sharp boundary between positive and negative cases of a vague predicate. One might for example be told by a usually reliable informant that 499,999 is the last small number. All other things being equal, one might thereby have a positive degree of justification for believing that 499,999 is the last small number—at least in the sense that, in so believing, one would be doing epistemically better than believing that 499,999 is the last small number by sheer chance (for example, by choosing at random from a list the proposition that 499,999 is the last small number). Still, one is not warranted in so believing—given the real import of the available evidence, one is epistemically criticizable for believing that 499,999 is the last small number.⁶

Higher-Order Vagueness. The fifth and last phenomenon of the vagueness of a predicate relates to all the previous ones and is the manifestation of all

⁶Note the contrast between the genuine phenomenon that we are ignorant of a sharp boundary between the small numbers and the non-small numbers and the alleged phenomenon that we are ignorant as to whether a borderline small number is small (or not small). While, under some natural assumptions, ignorance with respect to each and every borderline case would indeed entail ignorance of a sharp boundary (see section 3.2.1), the converse entailment does not hold, since, letting i be the sharp boundary for smallness, it may be the case that, in a certain situation, a subject s_0 comes to know that i is small and, in another situation, a subject s_1 comes to know that $i + 1$ is not small without s_0 and s_1 ’s ever being able to pool together these pieces of knowledge in order to come to know that i is the sharp boundary for smallness.

the previous phenomena by the vocabulary needed to describe them. ‘Sorites susceptible’, ‘indefinite’, ‘borderline case’ and ‘ignorant of a sharp boundary’ are all such that they seemingly fail to draw a sharp boundary between their positive and negative cases, such that they fail to draw a definite sharp boundary between their positive and negative cases, such that they possibly present borderline cases and such that we are presently ignorant of a sharp boundary between their positive and negative cases.

The literature has devoted special attention to the following borderline-case version of this general phenomenon. The distinction between small and non-small numbers is vague and hence there are borderline small numbers. But the distinction between borderline small numbers and not borderline small numbers is itself vague and hence there are borderline borderline small numbers. But the distinction between borderline borderline small numbers and not borderline borderline small numbers is itself vague and hence there are borderline borderline borderline small numbers etc.⁷ Formal constructions based on this and similar informal iterative principles have been developed and can be extended well into the transfinite (see e.g. Fine [1975], pp. 287–98; Williamson [1999]).

At first, higher-order vagueness might look like a perplexing phenomenon, but, rather than indicating an implicit complexity and hierarchical structure in our understanding of what it is for a predicate to be vague, it is probably just due to the *omni-pervasiveness* of vagueness in natural language (and so in particular in words like ‘sorites susceptible’, ‘indefinite’, ‘borderline case’ and ‘ignorant of a sharp boundary’). This is not to deny, of course, that such an omni-pervasiveness stands itself in need of an explanation (see chapter 3 for a discussion of some of the most problematic features of higher-order vagueness).

⁷With respect to borderline cases, Wright [2007b] proposes a different understanding of what higher-order vagueness should consist in. In this essay, I will stick to the traditional understanding just sketched.

1.4 The Theories of Vagueness

Given what the phenomena of vagueness are, I would like to contrast two different approaches to contemporary theorizing about vagueness. On the one hand, the *dominant approach* takes as basic the phenomenon of borderlineness, interpreting it as strong borderlineness. Consequently, this approach sees as its main theoretical tasks the explication of the notion of being borderline (or, equivalently, of being definite) and the explanation of the other phenomena of vagueness (especially, sorites susceptibility, ignorance and higher-order vagueness) by means of the more basic phenomenon of borderlineness.

Whatever its other merits, it is worth noting that such an approach is in a constant danger of depriving vagueness of its status of an interesting “natural” property of linguistic expressions, since more often than not the explication of the notion of being borderline on offer does not discriminate between *vagueness-specific borderlineness* (that is, borderlineness which is accompanied by the other phenomena of vagueness) and a more general kind of *indeterminacy* (either semantic, or ontic, or epistemic, or psychological, or primitive) which might affect linguistic expressions (think for example of the indeterminacy affecting the question whether a formerly two-headed man with exactly one of his heads cut off has been beheaded).⁸

Vagueness is recovered by adding a (usually not very convincing) story as

⁸This is an instance of what Alston [1964], pp. 87–90 calls ‘*combinatorial vagueness*’, distinguishing it from what he calls ‘*degree vagueness*’, which is vagueness in our sense. Roughly, combinatorial vagueness arises from indeterminacy as to whether a certain pattern of scoring along different dimensions relevant to the application of a predicate is sufficient for its positive (or negative) application. Not all indeterminacy is either degree vagueness or combinatorial vagueness: think for example of the indeterminacy affecting the question whether an electron released at a certain time by a helium ion after being in a superposition state is identical with the electron captured by the ion at an earlier time (see Lowe [1994]) or, letting a and b be the two complex square roots of -1 , think of the indeterminacy affecting the question whether $i = a$ (see Brandom [1996]).

to why, by some pragmatic or psychological mechanism, the relevant predicates suffering from indeterminacy come to be sorites susceptible. Vagueness turns out then to be a complex conjunctive property, one of whose conjuncts is exemplified also by predicates like ‘beheaded’: genuinely vague predicates like ‘bald’, ‘heap’ and ‘red’ are distinguished from it only insofar as they also happen to satisfy a second (rather uninteresting) conjunct to the effect of there being a pragmatic or psychological mechanism which produces sorites susceptibility. Moreover, it also turns out that the factors usually supposed to trigger such a mechanism (arbitrariness and unknowability of any boundary that the predicate were to possess) are just as well present for predicates like ‘ ξ bricks would break a camel’s back and no one will ever know that ξ bricks would break a camel’s back’, which for all intents and purposes can count as precise in this context (and, in particular, as not sorites susceptible), but whose sharp boundaries are just as arbitrary and unknowable as those that would be possessed by vague predicates.

On the other hand, the *traditional approach* takes as basic the phenomenon of sorites susceptibility. Consequently, this approach sees as its main theoretical tasks the explanation of this phenomenon, an account of its relation with sorites paradoxes and the explanation of the other phenomena of vagueness (especially, borderlineness, ignorance and higher-order vagueness) by means of the more basic phenomenon of sorites susceptibility. I side squarely with the traditional approach. The kind of theory falling within it which I should like to defend in this essay is what may aptly be called ‘a *naive theory of vagueness*’. A naive theory of vagueness holds that the key to the *nature* of vagueness—to what vagueness *consists in*—lies in taking sorites susceptibility at face value: very roughly, the vagueness of a predicate consists in its failure to draw a sharp boundary between its positive and negative cases—in what, following Wright [1975], pp. 333–4, we may call ‘its *tolerance*’ (sorites susceptibility can then be glossed as the appearance of tolerance).⁹

⁹I will give some references in fn 13 as regards modern-day naive theorists. What

Of course, this characterization needs many refinements. In its talk of ‘boundary between’, it presupposes suitable *orderings* of the range of significance of a predicate, which indeed do seem to be always available for vague predicates. The characterization should also be understood as presupposing that on at least some orderings where there is no sharp boundary there are *both positive and negative* cases (so that nihilist views as espoused in Unger [1979a]; Unger [1979b]; Unger [1979c]; Unger [1980], reference-failure views as espoused in Wheeler [1975]; Wheeler [1979]; Braun and Sider [2007] and incoherentist views as espoused in Dummett [1975a]; Dummett [1979] (on one possible understanding of this author’s contention); Rolf [1981], pp. 116–49; Rolf [1984] are ruled out from naivety) and as presupposing that such positive and negative cases are “positive-*only*” and “negative-*only*” respectively (so that dialetheist views as briefly mooted in Priest [2003], p. 9 and incoherentist views as espoused in Dummett [1975a]; Dummett [1979] (on another possible understanding of this author’s contention) are ruled out from naivety). Possibly with the exception of certain *recherché* cases to be discussed shortly, the characterization should further be understood as presupposing the correctness of the inference from x ’s being a positive (negative) case (together with the relevant instance of tolerance) to anything very close to x in the relevant ordering also being a positive (negative) case—that is, as presupposing the validity of *modus ponendo ponens* and of attendant arguments like *modus ponendo tollens* (so that all the many different views according to which these arguments are invalid, as espoused in Peacocke [1981]; Hyde and

other kinds of theories or positions belong to the traditional approach? Unsurprisingly, most ancient positions do (see Masson-Oursel [1912], pp. 820–4; Moline [1969]; Barnes [1982]; Burnyeat [1982]; Sillitti [1984]; Mignucci [1993]; Bobzien [2002] for presentation of some of these). Many modern nihilist, reference-failure and incoherentist theories do (the clearest cases are Dummett [1975a]; Dummett [1979]; Unger [1979a]; Unger [1979b]; Unger [1979c]; Unger [1980]; Rolf [1981], pp. 116–49; Rolf [1984]; Eklund [2001]; Eklund [2005]). Some theories exploiting the formalism of fuzzy logic do (a clear case is Priest [1998]). Most prominently nowadays, many contextualist views do (the clearest cases are Raffman [1994]; Raffman [1996]; Fara [2000]).

Sylvan [1993]; Hyde [1997]; Priest [1998]; Varzi [2000], are ruled out from naivety). The characterization should finally be understood as presupposing that there only need not to exist *one* sharp boundary between positive and negative cases (so that ‘ ξ is in her early thirties’ does count as vague, see Weatherson [2005b], p. 2).

Even when such and similar refinements are in place, I am under no illusion that anomalies may still arise requiring yet further qualifications. Just to give an example, it might be the case that a certain predicate ‘ F ’ is such that, on a relevant ordering from F s to non- F s:

- It is in some sense unsettled whether a certain object x is F (where ‘unsettled’ is used of course in such a way as not to imply¹⁰ attribution of strong borderlineness);
- It is intuitively correct to deny the conditional that, if x is F , so is some object following x in the ordering;
- It seems that no sharp boundary is drawn by ‘ F ’ before x .

‘ ξ is not very late and ξ is at most 10 minutes late’ might be such a predicate (I am extrapolating here from an example of Weatherson [2006], pp. 4–5 what seems to me really problematic for the naive theory). Intuitively, if x is F , ‘ F ’ is not vague and, if x is not F , ‘ F ’ is vague, and so it is unsettled whether ‘ F ’ is vague. However, it is not the case that, if x is F , so is some object following x in the ordering and hence we would be forced to accept that ‘ F ’ is not vague after all.

One reply would be to restrict the naive theory of vagueness to “basic cases” of vague predicates and to explain the vagueness of other predicates as arising in virtue of the vagueness of some basic predicates occurring in

¹⁰Following a rather established usage, I will throughout use ‘implication’ and its like to denote the operation expressed by ‘If φ then ψ ’, ‘entailment’ and its like to denote the converse of logical consequence (see section 5.1).

them (or in the meaning explanations which introduce them).¹¹ If one is sceptic as to whether a workable notion of a “basic vague predicate” can be made out and wants to preserve the intuition that it is unsettled whether ‘*F*’ is vague, I suggest that one should drop the fusion version of tolerance from the naive theory (of course, if one does not recognize any admissible sense of ‘and’ which does not collapse on extensional conjunction, the problem does not arise in the first place, since the relevant negated extensional conjunction is unsettled). This would be of course not to deny that a naive theorist can still hold that satisfaction of the fusion version of tolerance is a highly reliable—though not infallible—guide to vagueness. I would offer the same two options in the case of predicates which are intuitively vague but whose meaning is such that at most one object belongs to their extension and so such that the fusion version of tolerance arguably fails for them (like ‘ ξ is identical to Kilimanjaro’, ‘ ξ is the best football player ever’, ‘ ξ is my favourite dish’). Analogous remarks apply to sorites susceptibility as a relatively uncontroversial necessary condition for vagueness.

It is easy to see how a naive theory of vagueness, together with the assumption that its truth is available to vague-concepts users, can explain the various phenomena of vagueness:

- A vague predicate seems to fail to draw a sharp boundary between its positive and negative cases just because it indeed fails to do so and this truth is available to vague-concepts users;
- Factivity of ‘definitely’ ensures that, if a vague predicate fails to draw a sharp boundary between its positive and negative cases, it also fails to draw a definite sharp boundary between them;
- If, for every case in a series, either a positive or a negative ‘small’-judgement about it were warranted, given that it is usually not vague which judgement is given about a particular case, it would follow, together with some natural assumptions about which judgements can be

¹¹Thanks to Sebastiano Moruzzi for this suggestion.

warranted, that, in a series starting with positive cases, there usually is a last case which is warrantably judged to be positive followed by a case which is warrantably judged to be negative. The judge's warrant to issue both these judgements is defeated by the availability to her, a vague-concepts user, of the truth that there is no sharp boundary in the series between positive and negative cases (in the case of knowledge and other factive epistemic properties, an even more straightforward explanation is available, since factivity of 'know' ensures that, if a vague predicate fails to draw a sharp boundary between its positive and negative cases, no object is known to be a sharp boundary between them);¹²

- The availability to a vague-concepts user of the truth that there is no sharp boundary between positive and negative cases defeats her warrant for identifying any object as the sharp boundary between positive and negative cases (again, in the case of knowledge and other factive epistemic properties, an even more straightforward explanation is available on the lines just sketched for borderlineness);
- An account of the sources of lack of sharp boundaries (as the one offered in chapter 2) will show them to be general enough as to predict that expressions like 'sorites susceptible', 'indefinite', 'borderline case' and 'ignorant of a sharp boundary' are also affected by them.

This explanatory power of the naive theory of vagueness is admirable. To my mind, together with its undeniable apparent self-evidence, it makes it the theory which must be defeated if it is not to be accepted. In the rest of this essay, I will be concerned part with strengthening even more the theory's credentials, both by offering detailed arguments for its main claim (chapter 2) and by working out its explanation of borderlineness (chapter 6); part with defending it in the face of the paradoxes of vagueness, by showing

¹²I'm skipping over important subtleties here: see chapter 6 for an elaborate and specific development of this kind of explanation.

how sorites paradoxes do not really tell specifically against it (chapter 3), by providing both a suitable logic which makes the theory consistent (chapter 4) and a general philosophical commentary to the main innovative feature of the logic (chapter 5) and by addressing a non-inferential paradox (chapter 6). Together with the apparent self-evidence and the explanatory fruitfulness just reviewed, I will regard success in these tasks to constitute a powerful cumulative argument in favour of the naive theory.¹³

¹³In modern times, the naive theory of vagueness has had very few defenders, probably because of the considerable technical and conceptual difficulties involved in stabilizing it. Ziff [1974], pp. 530–1; Ziff [1984], pp. 140–4 constitute lively, if brisk, endorsements of the theory, echoing many pre-theoretical intuitions elicited by vagueness and the sorites paradox. Weiss [1976] presents a carefully crafted framework aimed at accommodating the theory: unfortunately, the framework is arguably misconceived in some of its key aspects and, in any event, lacks the required generality, dealing as it does only with induction versions of the sorites paradox. The same lack of generality affects the remarks in favour of the theory in Vopěnka [1979], pp. 33–4, whose controversial mathematical apparatus, while providing an explanation of the failure of induction on vague properties, does nothing to explain what goes wrong in more elementary versions of the sorites paradox (as those studied in chapter 4), which make use only of simple rules of inference such as *modus ponens* (classical logic goes unchallenged throughout Vopěnka’s monograph). Kamp [1981] is a very rich and stimulating discussion of several importantly different context-dependent consequence relations: the spirit of the proposal is certainly aligned with the naive theory, accepting that there is no sharp boundary between positive and negative cases, even though one prominent family of consequence relations in Kamp’s study fail to validate *modus ponens* (and attendant arguments) and his discussion of another prominent family is unfortunately marred by an ill-motivated constraint on their model theory. Parikh [1983] is a most insightful discussion, to which this essay is greatly indebted: the final pages (which should be read along with their technical companion Parikh [1971]) contain a very brief sketch of how the naive theory should be developed and have been a source of inspiration for my own account. Although I disagree with much of its details, Rasmussen [1986] also develops some interesting lines in defence of the theory. Finally, Gaifman [2007] constitutes a significant recent defence of the theory: his arguments in favour of tolerance are however not very convincing and the consistency of the theory is achieved only at the cost of highly counterintuitive contextual domain restrictions.

1.5 Plan

As I have just anticipated, the plan for the rest of this essay will be as follows. Chapter 2 provides a battery of arguments to the effect that tolerance is indeed a necessary condition for some of our concepts to serve in the achievement of certain important theoretical and practical purposes which have traditionally been associated with vagueness. Chapter 3 shows how, once higher-order vagueness is taken seriously, higher-order sorites paradoxes fully analogous to standard sorites paradoxes can be developed by making use of borderlineness principles instead of tolerance principles. Chapter 4 develops a family of logics which, by placing principled restrictions on the transitivity of the consequence relation, invalidate the sorites paradox and make the naive theory consistent. Chapter 5 discusses some of the main issues involved in making sense of reasoning in a non-transitive logic. Chapter 6 offers a solution to another problem widely discussed in the contemporary debate on vagueness and uses it to explain the borderlineness phenomenon without recourse to strong borderline cases.

Before concluding this introduction, a word on at least some of the issues that, for reasons of space, will *not* be discussed in this essay, but would certainly have to be taken into account in a fuller treatment of the topic. As I have already said, I have not developed another important part of my views on vagueness, namely the rejection of the existence of strong borderline cases, since, while this doctrine certainly chimes with the traditional approach and in particular with a naive theory of vagueness, it is strictly speaking not required by them. Nor have I addressed the many interesting issues arising at the interface with metaphysics: in particular, I have not considered the vexed question as to whether vagueness is a linguistic or ontic phenomenon and I have not tackled a metaphysical problem which vagueness appears to give rise to (the well-known problem of the many), even though (as it might transpire from some of the remarks in chapter 2 and in section 1.4 respectively) I do have my own views on both these issues. Nor have I been able to go into

the complex range of issues specifically pertaining to so-called “phenomenal sorites”, in particular into the question of the nature and properties of the indiscriminability relation. Nor have I tried to determine what bearing the technical and conceptual apparatus developed especially in chapters 4 and 5 respectively may have on the debate on strict finitism. Finally, save for some cursory remarks, I have also quite generally not discussed in any decent detail extant alternative approaches to the problems I am concerned with, preferring to focus instead on the articulation of the positive arguments and considerations in favour of my views.

Chapter 2

Seconde Naïveté

2.1 Introduction and Overview

There is a very general schematic principle that an 1ary predicate ‘ F ’ might satisfy, principle which would arguably reveal to us something very interesting about ‘ F ’ (and about the F s themselves). If R is a particular (reflexive, symmetric but possibly non-transitive) relation of closeness along a dimension of comparison relevant for being F , the schematic principle can be expressed as follows:

- (N) For every x, y such that $x R y$ and y is otherwise just the way x is (at least as far as ways relevant for being F are concerned), it is not the case that x is F and y is not F .

For example, plausible instances of (N) are:

- (N^{small}) For every numbers x, y such that [either $x = y + 1$ or $x = y$ or $x = y - 1$] and y is otherwise just the way x is (at least as far as ways relevant for being small are concerned), it is not the case that x is small and y is not small;

(N^{high}) For every mountains x, y such that x is at most one nanometre either higher or lower than y and y is otherwise just the way x is (at least as far as ways relevant for being high are concerned), it is not the case that x is high and y is not high;

(N^{rich}) For every persons x, y such that x has at most one penny more or less than y and y is otherwise just the way x is (at least as far as ways relevant for being rich are concerned), it is not the case that x is rich and y is not rich.

Before proceeding further, three important remarks concerning (N) and my understanding of it should be made. Firstly, the qualification ‘ y is otherwise just the way x is (at least as far as ways relevant for being F are concerned)’ is intended to take care of the phenomenon of *multi-dimensionality* so frequently occurring with vague predicates—that is, of the fact that the correctness of the application of many a vague predicate to a particular item depends on the item’s location on at least two distinct dimensions of comparison. For example, the correctness of an application of ‘bald’ to a man depends not only on the number of hairs on his scalp, but also on their distribution, density, thickness etc. Because of multi-dimensionality, the simple no-sharp-boundaries principle that it is not the case that [a man is bald and a man with just one more hair on his scalp is not],¹ is, strictly speaking, false: a man with a sufficient number i of hairs, widely distributed, homogeneously dense and appropriately thick may well count as non-bald, whereas a man with $i - 1$ hairs so poorly distributed as to cover only a square centimetre of his scalp, so heterogeneously dense as to leave a hairless circle in the middle, so thin as to be invisible will count as bald. Therefore, simple no-sharp-boundaries principles like the one just mentioned apply straightforwardly only to uni-dimensional predicates—in the case of multi-dimensional ones, a no-sharp-boundaries principle applies on a particular dimension of comparison only under the assumption that the values of the other dimen-

¹Throughout, I will use square brackets to disambiguate scope in English.

sions are held constant. Hence the need for the qualification ‘ y is otherwise just the way x is (at least as far as ways relevant for being F are concerned)’.

Secondly, (N) has almost universally been thought to be inconsistent with there being both positive and negative cases of F ness which (possibly) differ with respect to the R -relevant dimension of comparison but are otherwise identical (at least as far as ways relevant for being F are concerned). The inconsistency is supposed to be revealed by the kind of argument long known as ‘sorites paradox’. I will show in chapter 4 that a satisfactory weakening of the logic exists in which principles of the form of (N) are consistent with the existence of the relevant positive and negative cases. In the following, I will argue that we do use some of our predicates so as to conform to the relevant instances of (N)—or, at any rate, that such a use is a necessary condition for achieving some important theoretical and practical purposes. As a consequence of the logical point just mentioned, I will neither accept nor consider the overall objection to my arguments to the effect that they must be wrong since our use of the predicates in question is consistent—or since no important theoretical or practical purpose requires an inconsistent theory—whilst (N), together with the aforementioned assumptions, would breed inconsistency. It just needn’t be so.

Thirdly, what I propose to do in this chapter is to investigate whether giving up (N) is a viable option in our theoretical and practical life. In this regard, I should note that I’m actually much less interested in the categorical, empirical and sociological claim that we in effect use some of our predicates so as to conform to (N) than I am in the hypothetical, philosophical and normative claim that such a use would achieve important theoretical and practical purposes—or, at least, is a necessary condition for achieving such purposes. I do find the arguments to follow compelling also with respect to our actual use and will therefore argue for the stronger conclusion, but I will ultimately rest content with simply showing that such a use is fully intelligible, highly valuable and hardly dispensable. Once its fine architecture and details have been brought out, I hope it will be clear that, even if it is

not our actual use, it is something we should strive to incorporate into our conceptual repertoire.

Clearly, principles of the form of (N) bear some interesting relation to vagueness. Indeed, the naive theory of vagueness holds, very roughly, that satisfaction of some such principle is what vagueness of a predicate ultimately consists in—what the *nature* of vagueness is (see section 1.4). Suitably developed and refined, this is the theory I aim to defend in this essay. Admittedly, the vindication of such a theory is no simple task (important aspects of it have been and will be considered throughout this essay), but a crucial part of the dialectic against its rivals consists in exposing the high costs of giving up the connection between a predicate's vagueness and the satisfaction of some principle of the form of (N).

From the point of view of the naive theory of vagueness, such satisfaction is what vagueness consists in, and so what these principles are for (the otherwise impossible achievement of important theoretical and practical purposes to be presently described) is what vagueness is for. In denying that such satisfaction is what vagueness consists in—indeed, in denying that such satisfaction is ever so much a necessary condition for vagueness—the rivals of the naive theory commit themselves to denying that vagueness is conducive to these important theoretical and practical purposes. Moreover, since it is very plausible that vagueness would be so conducive if anything were, the rivals commit themselves to denying that such purposes can possibly be achieved. Therefore, the more valuable these purposes can be shown to be, the less appealing the rivals will appear. This specific part of the dialectic is what this chapter tries to accomplish, and, granting the success of the overarching project, what this part would then amount to is an exhibition of the grounds of the vagueness of a predicate—of what the *sources* of vagueness are. Let me stress that, even though I will identify several such sources, the discussion will by no means be meant to be exhaustive. Other sources of this complex phenomenon wait to be uncovered.

Before embarking in the details of this part of the dialectic, let me stress that it is on my view akin to the one involving the naive theory of *truth*. Even though, contrary to a highly suggestive line of thought put forward by some prominent commentators (McGee [1991]; Tappenden [1993]; Soames [1999]; Field [2003b]), I don't think that the problems concerning absolute semantic notions and those concerning vagueness have a common root, I do think that, just as a strong case can be made that nothing falling short of the naive theory of truth can do justice to our use of the concept of truth (and related semantic concepts such as reference, denotation, satisfaction etc.) as picking out a universal property of representational correctness (or, if you prefer, as fulfilling the function of a universal device of disquotation), an at least equally strong case can be made that nothing falling short of the naive theory of vagueness can do justice to great a many features of our use of vague concepts. This case will be set out in the following. Indeed, as will be seen, in the case of vague concepts, as opposed to the case of truth (and related semantic concepts), this strategy can be developed with respect to significantly different and apparently independent features of our use of such concepts, which partly explains why sometimes satisfaction of principles of the form of (N) will only be argued to be a necessary condition for a particular feature of use to achieve its purpose.

I also happen to think that seeing our way through the paradoxes which seem to jeopardize these theories (respectively, the semantic paradoxes and the sorites paradox) requires a deep rethinking of aspects naturally associated with the notion of logical consequence (though a different aspect is concerned in each case). As far as the naive theory of vagueness is concerned, this issue will be taken up in chapter 5.

Given this logical heterodoxy (which will be advocated and developed in chapters 4, 5), a last word is owed about the logic used in the informal deductive arguments throughout this essay. All these arguments are, I hope, classically valid. They all only appeal, I hope, to what are in context intuitively acceptable principles (there might be some exceptions involving

e.g. use of the law of excluded middle—where this is the case, I usually try to indicate how the argument may be modified in order to reach similar conclusions). Some of them are certainly not formally valid according to any of the logics developed in chapter 4. This presents us with a new instance of the well-known problem of, to put it somewhat roughly, recovering in context a logic stronger than that to which one has committed oneself to.

As will emerge in chapter 4, the only fault I find with classical logic as far as vagueness is concerned is that, by sorites reasoning, it allows us to use intuitively correct tolerance principles to go down the slippery slopes associated with vague predicates. I thus regard a classically valid argument as unacceptable (if and) only if it is soritical (the notion could be made more precise, but I trust that we already have a workable understanding of which arguments are soritical and will leave it at that in this essay). The question becomes then how to justify, in the face of the proposed weakening of the logic, the classically valid arguments which, although formally invalid (according to the proposed weakening), are not soritical (call these ‘the *good classical arguments*’).

There are different strategies for doing this. One strategy would be to accept *extra premises* which, when added to the premises of a good classical argument, transform it into a formally valid argument. The extra premises would be accepted on the grounds that they are true given that the original good classical argument is not soritical. It is actually not straightforward to implement this strategy in the case of the particular weakening of the logic proposed in chapter 4—this would require the addition of new operators to the language in order to express the needed extra premises, something which I will not try to do in this essay. For this and other reasons, I regard this strategy as at best very unnatural in the present case.

Another strategy, which in other cases would be barely distinguishable from the first one, is to accept that classical logic is *valid in the context* represented by a good classical argument, so that the argument turns out to be after all formally valid. Classical logic would be accepted as valid in the

context represented by a good classical argument on the grounds that the argument is not soritical. I regard this strategy as violating the universality implicit in the very formality of formal validity, and at best as a radically misconceived attempt at pursuing what is indeed a different, third strategy.

I myself would prefer this third strategy, which accepts good classical arguments as *materially*, even though *not formally*, valid (see Read [1994] for an illuminating discussion of the distinction between formal and material validity). A materially valid argument is, roughly, an argument which is valid at least partly in virtue of the occurrences in it of certain *non-logical*² expressions (see section 5.1). For example, the argument ‘There is nothing in the bottle. Therefore, the bottle is empty’, even though formally invalid (since we may assume that the most specific argument form it instantiates is ‘There is no x such that x *R*s a . Therefore, a is F ’), is intuitively valid in virtue of the occurrences in it of the expressions ‘there is nothing’, ‘in’ and ‘empty’.³ Of course, it is a major task in the philosophy of logic to specify the grounds for this intuitive judgement of validity, a task which I will not attempt to undertake in this essay. Here, I will rest content with pointing out, on behalf of the friends of material validity, that it seems very plausible that, whatever story is ultimately told about the grounds of the validity of ‘There is nothing in the bottle. Therefore, there is nothing in the bottle or snow is white’ (a story which will presumably involve the logical expression ‘or’ and talk about either some form of deducibility or about some form of truth preservation), it will also be possible to tell an analogous story about the grounds of material validity (that is, a story which will presumably involve the relevant non-logical expressions and talk about either some form

²Throughout, I assume a rough-and-ready understanding of logicity. Nothing will hinge on the fine details of this complex notion (see Gómez-Torrente [2002a] for a recent critical introduction to the issue).

³Note that, while the first strategy too could be applied in this case (by supplying the extra premise ‘For every x , if there is nothing in x , then x is empty’), the second strategy is a non-starter, since the argument form in question is not even classically valid and, to put it somewhat roughly, no supra-classical logic is formal.

of deducibility or about some form of truth preservation).

I take it to be a fact of life that some arguments strike us as correct not in virtue merely of their *abstract logical form*, but at least partly in virtue of their *specific subject matter*, which is crucially determined by the occurrences in them of non-logical expressions. The notion of material validity enables us then to take these appearances at face value (what the first strategy can never do and, as argued in fn 3, the second strategy cannot always do). It thereby entitles us to *reason* (i.e. draw inferences) in ways which, albeit not sanctioned by the peculiar universality attaching to logical formality, are intuitively correct at least partly in virtue of substantial non-logical details of the subject matter at hand. It makes our understanding of ‘in’ and ‘empty’ a genuine source of reason no less than our understanding of ‘or’.

To come to the application of the formal/material distinction employed by the present strategy, good classical arguments are accepted as materially valid on the grounds that examination of the non-logical vocabulary occurring in them reveals that they are not soritical. Anticipating a little, let us consider as an example the following argument, given in section 2.2.3:

[S]uppose that x is not close enough [...] Then, by (WO), on at least some occasion it can be known by casual observation that x is not close enough, and so, by (SI), on no occasion can it be known by casual observation that y is close enough. But then, by (WO), y is not close enough.

For present purposes, the argument can informally be regimented as ‘ x is not close enough; If x is not close enough, then on at least some occasion it can be known by casual observation that x is not close enough; If on at least some occasion it can be known by casual observation that x is not close enough, then on no occasion can it be known by casual observation that y is close enough; If on no occasion can it be known by casual observation that y is close enough, then y is not close enough. Therefore, y is not close

enough’, which instantiates the argument form ‘ P_0 ; If P_0 , then P_1 ; If P_1 , then P_2 ; If P_2 , then P_3 . Therefore, P_3 ’. According to all the logics developed in chapter 4, this argument form is invalid, and so the argument itself which instantiates it is not formally valid. Indeed, it is bound not to be so, since the same argument form is instantiated by the argument ‘0 is small; If 0 is small, so is 1; If 1 is small, so is 2; If 2 is small, so is 3. Therefore, 3 is small’, which no naive theory of vagueness should be willing to recognize as valid.

Still, I accept the conclusion of the argument in section 2.2.3. Indeed, I accept it on the grounds of the argument provided, and I do so because I recognize that argument to be (materially) valid, even though it is not formally so. I recognize the argument to be materially valid at least partly because of the occurrences in it of certain non-logical expressions. In this and other cases of interest in the present context, it is not so much that, intuitively speaking, there is something “positive” about the occurrences of the non-logical expressions (‘close enough’, ‘on’, ‘occasion’, ‘can’, ‘know’, ‘casual observation’) which contribute to the validity of the argument (as was the case, on the contrary, for the previous argument involving ‘in’ and ‘empty’)—it is rather that there is nothing “negative” about them, nothing which makes the argument soritical and hence objectionable. From the point of view of the logics developed in chapter 4, the formal invalidity of the argument shows that the occurrences in it of the logical expression ‘If φ , then ψ ’ do not suffice to ensure the nexus of consequence between the premises and the conclusion—such is the lesson of the sorites paradox with regard to the logical power of conditional chains. What does so suffice is rather the *combination* of those occurrences with the occurrences of non-logical expressions ruling out the soriticality of the argument.^{4,5}

⁴Thanks to Stephen Read for pressing me to think hard about my employment of the formal/material distinction in developing this strategy.

⁵In closing this digression, let me make clear that, despite my preference for the third strategy just considered, I don’t take myself to have adjudicated here on this difficult issue in the philosophy of logic. The reader who is attracted by the general outlook of this essay is invited to reconstruct the good classical arguments contained in it according to her

The rest of the chapter is organized as follows. Section 2 shows what (N) allows us to think about the world in terms of classifications of objects which can sort them into interesting theoretical, practical and emotional kinds, be flexible and be achievable by relying only on casually available evidence. Section 3 shows what (N) allows us to experience of the world in terms of seamless changes and appearances (the stark distinction between the two sections is partly dictated by presentational needs: as will be seen, some arguments really pertain to both sides). Section 4 draws the conclusions which follow from the specific arguments given for what, according to the naive theory of vagueness, the sources of vagueness are.

2.2 Thoughts Requiring the Absence of Sharp Boundaries

2.2.1 Irrelevant Differences

In this first section, I want to expand on and generalize a suggestive line of thought first put forward by Wright [1975] (cf Wright [1976]). In considering the high plausibility of principles of the form of (N) for age nouns like ‘child’, ‘adolescent’, ‘adult’ etc., Wright remarked that the classifications induced by such nouns “are of substantial social importance in terms of what we may appropriately expect from, and of, persons who exemplify them” (Wright [1975], p. 336). He then observed that, on the one hand, “[i]t would be irrational and unfair to base substantial distinctions of right and duty on marginal – or even non-existent – such differences” and that, on the other hand, “[o]nly if a *substantial* change is involved in the transition from childhood to adolescence can we appeal to this transition to explain substantial alterations in patterns of behaviour” (Wright [1975], p. 337). On these grounds, he concluded that, for such age nouns, “very small differences cannot be permitted to generate

favoured strategy.

doubt about their application without correspondingly coming to be associated with a burden of moral and explanatory distinctions which they are too slight to convey” (Wright [1975], p. 337).

I think we can extract from this the following quite general pattern of requirements of respectively *important* and *unimportant differences* that we sometimes impose on being *F*. That is, sometimes:

- (ID) We attach great importance to being *F* rather than falling in some sense short of being such;⁶
- (UD) We do not attach any great importance to minute differences with respect to the *R*-relevant dimension of comparison.

Before proceeding to apply this pattern beyond age nouns, it is important to clarify the function in this dialectical context of the phrase ‘falling in some sense short of being *F*’ and its like. In the diverse arena of contemporary theories of vagueness, there are many different ways in which *x* can *fall short of* being *F* without thereby being guaranteed to be *not F*. The range of the alternative honorific candidates is wide and well-known:

- Negation of the negation of the proposition that *x* is *F*, accompanying rejection⁷ or even negation of the proposition that *x* is *F*;
- Negation of the proposition that *x* is *unF* (where ‘*unF*’ is the proximate contrary of ‘*F*’), accompanying negation of the proposition that *x* is *F*;

⁶Focus on the (possibly) weaker requirement that we attach great importance to being *F* rather than *not* being such and consequent failure to pay due heed to the stronger (ID) seem to me to flaw the considerations advanced in Sainsbury [1989], pp. 38–9; Sainsbury [1995], p. 28.

⁷Throughout, *rejection* (along with *denial*, its speech-act manifestation) will be understood as a *primitive* attitude on a par with acceptance—in particular, it will not be presupposed that it implies or is implied by acceptance of the corresponding negation (see Parsons [1984]; Smiley [1996]; Tappenden [1999]; Rumfitt [2000]; Field [2003b]; Priest [2006b], pp. 103–15 for various arguments in favour of positing this distinctive attitude).

- Negation of the proposition that x is either F or not F ;
- Rejection of the negation of the proposition that x is F , accompanying rejection that x is F ;
- Rejection that x is either F or not F ;
- Negation of the proposition that it is true that x is not F , accompanying negation of the proposition that it is true that x is F (or negation of the proposition that ' x is not F ' is true, accompanying negation of the proposition that ' x is F ' is true), and higher-order variations thereof;
- Negation of the proposition that it is definitely (or determinately, or clearly etc.) the case that x is not F , accompanying negation of the proposition that it is definitely (or determinately, or clearly etc.) the case that x is F , and higher-order variations thereof;
- Acceptance that x is F only to some intermediary degree

and many others.

Relatedly, even better established are attempts at formulating principles which, while weaker than those licensed by (N), still try to preserve some of the intuitive force behind them, mostly by also allowing some strengthening of the original supposition that x is F . Some such principles are:

- If x is definitely (or determinately, or clearly etc.) F , then y is not definitely (or not determinately, or not clearly etc.) not F (and their truth-theoretic analogues);
- If x is definitely (or determinately, or clearly etc.) F , then y is F (and their truth-theoretic analogues);
- It is not both acceptable that x is F and rejectable that y is F ;
- If x is F to a certain degree, y is F at least to a not significantly smaller degree.

Even setting aside the question of their (doubtful) dialectical efficacy in preserving the spirit of (N) (without its alleged paradoxical consequences), such principles will be irrelevant here, our assumption being just the plain one that x is F , and our question being what follows from that with respect to y 's F ness.

To conclude this clarification, I will henceforth use the catch-all phrase 'falling in some sense short of being F ' and its like for every alternative honorific candidate to being F . As a first approximation, the class of all such alternative honorific candidates can usefully be delimited by the condition that exemplifying the candidate is inconsistent with exemplifying F ness.⁸ We are now ready to expand on (ID) and (UD) following through their consequences for the truth of principles of the form of (N) and also ready to generalize their range of application to cases beyond age nouns. We will accomplish both tasks in one go by considering cases of predicates belonging to the realms of theory (both scientific and ordinary), action and feeling.

Theory. Suppose that x is a dog. Suppose that y differs from x at most for the fact that one of y 's atoms is in a location which is within a nanometre distance from the location correlating to the location of the corresponding atom part of x . Then it should be correct to assert that y is a dog as well—*anything* falling in some sense short of this would seem to draw an *arbitrary* difference between x 's and y 's animal status based on a difference, such as the nanometrical displacement of a single atom, which we perceive to be fully *irrelevant* for dogs.

The reason for the perceived irrelevance seems clear enough. Purporting to pick out a biological kind, doghood, our concept of dog is supposed to individuate an objectively distinguished feature of nature. We aim at describing important biological facts by referring to instantiation of such a kind

⁸Note that, if e.g. an epistemic spin in terms of clarity is given to the notion of determinacy, then the relevant alternative honorific candidate is negation of the proposition that it is clearly the case that x is not F , accompanying negation of the proposition that it is clearly the case that x is F and acceptance that x is, after all, not F .

by a living being rather than instantiation of anything falling in some sense short of it.⁹ In its small and limited way, doghood is thus thought to “carve nature at its joints”. But there is no relevant biological joint to be carved in a nanometrical difference of one atom’s location: life just doesn’t go that deep. Thus, under the supposition that x is a dog, we are forced to reject any predication about y entailing that y falls in some sense short of being a dog, and so we are forced to accept that y is a dog. (To counter a likely rejoinder, note that rejection that y falls in some sense short of being a dog, accompanying rejection that y is a dog, was itself understood to be—in a typically reflexive way—one of the senses in which y can fall short of being a dog.) By *reductio*, the truth of the relevant instance of (N) follows.

Analogously to Wright’s discussion of age nouns, the irrelevance of nanometrical differences can be made to emerge also from a different angle. We may assume that, under certain circumstances, x reacted in a certain way to the ingestion of a certain pill *because* x is a dog. But, certainly, it is not the case that x reacted in that way because x ’s atoms are arranged in at most such and such distances from one another and it is not the case that the distance of even just two of them is a nanometre greater: in such a context, that x falls on one side rather than the other of such an exquisite distinction has no scientific explanatory interest, as opposed to its falling on one side rather than the other of the distinction between being a dog and falling in some sense short of being such. This situation can be usefully contrasted with clear cases of reduction of a higher-level property to a lower-level property (even when the explanation in question is an explanation of action): assuming that the fact that there was water in the glass is a good (even though partial) explanation of the fact that Nancy reached for the glass, so is the fact that there was H₂O in the glass. Thus, by *reductio*, the property of being a dog cannot be coextensive with any *sharply bounded* property (that is, a property which discriminates with the highest possible degree of preci-

⁹Throughout, I will often make use of kind- and property-talk. This is only for ease of exposition and should be considered as ultimately dispensable.

sion between objects which exemplify it and object which fall in some sense short of exemplifying it). Since it would be so coextensive if the relevant instance of (N) were not to be accepted, the relevant instance of (N) is to be accepted.¹⁰

In presupposing that, if (N) is not to be accepted, the distinction between being a dog and falling in some sense short of being such must be nanometrical, this argument presupposes an exhaustive bipartition on the relevant domain between dogs and objects falling in some sense short of being such, and such bipartition in turn presupposes the truth of the relevant instances of the law of excluded middle. But the argument can easily be restated without this assumption: even in a logical framework where the law of excluded middle fails (for various proposals on how to do this in the case of vagueness, see Burgess and Humberstone [1987]; Tappenden [1993]; Soames [1999]; Wright [2001]; Field [2003a]), it would be correct to say that there is no explanatory interest in the fact that *x* falls on one side rather than the other of a distinction which is denied to be *any coarser* than one determined by a nanometrical difference of one atom's location. (Note that, even by the lights of the just mentioned proposals on which the law of excluded middle fails, the universal closure of the relevant instance of (N) must be rejected, since, together with uncontroversial assumptions, it leads to contradiction.)

The point can be made in an equally forceful and yet tellingly slightly different way for concepts which do not purport to pick out natural kinds. Consider the concept of baldness. It too is entrenched in a sophisticated (folk) theory—about the physiological causes of baldness, the way bald persons look, the social import of baldness etc. But, on the one hand, none of these diverse aspects discriminate importantly between neighbouring numbers of hairs on a person's scalp: two causes such that one differs from the other only in causing the loss of just one more hair are not importantly different, one's look is not importantly altered by the addition of a single hair, nor

¹⁰I owe the inspiration for this argument from explanation to a remark put to me by Stephen Schiffer.

is one's social impact. On the other hand, baldness is supposed *always* to affect *importantly* these aspects. We expect any other state of a person's scalp falling in some sense short of being bald to be due to substantially different causes than those known to bring about baldness—otherwise, why worry about also preventing causes merely similar to the latter? We expect a bald person to look substantially different from anyone falling in some sense short of being bald—otherwise, why be relieved about one's look upon being told that one still falls in some sense short of being bald? We expect a bald person to have a substantially distinctive social impact—otherwise, why should uncertainty about one's baldness engender uncertainty as to how one will be received for the first time by one's partner's parents? It follows from this contrast that baldness should not be sensitive to one-hair differences or, worse, to a nanometrical difference of one atom's location: baldness too just doesn't go that deep. It too only tracks a *coarse* distinction, even if not a natural one. Thus, under the supposition that x is bald and y only has one hair more on his scalp, we are forced to reject any predication about y entailing that y falls in some sense short of being bald, and so we are forced to accept that y is bald. By *reductio*, the truth of the relevant instance of (N) follows.

Again, this irrelevance can be made to emerge also from a different angle. We may assume that Sally no longer goes out with x *because* x has become bald. But, certainly, it is not the case that Sally now behaves the way she does because x 's atoms are arranged in at most such and such distances from one another and it is not the case that the distance of even just two of them is a nanometre greater: in such a context, that x falls on one side rather than the other of such an exquisite distinction has no ordinary explanatory interest, as opposed to his falling on one side rather than the other of the distinction between being bald and falling in some sense short of being such. Thus, by *reductio*, the property of being bald cannot be coextensive with any sharply bounded property. Since it would be so coextensive if the relevant instance of (N) were not to be accepted, the relevant instance of (N) is to be

accepted.

Action. Suppose that x is a person. Suppose that y differs from x at most for the fact that one of y 's atoms is in a location which is within a nanometre distance from the location correlating to the location of the corresponding atom part of x . Then it should be correct to assert that y is a person as well—*anything* falling short of this would seem to draw an *invidious* difference between x 's and y 's personal status based on a difference, such as the nanometrical displacement of a single atom, which we perceive to be fully *irrelevant* for persons.

Again, the reason for the perceived irrelevance seems clear enough. Personhood is embedded in a rich web of commitments and entitlements, among which prominent are practical ones. Importantly different actions are licensed with respect to someone who can be said to be a person and someone who cannot be said to be such. And we simply haven't come up with the concept of a person to find ourselves forced to discriminate in such important ways between two fellow beings differing only in such unimportant respects.¹¹ Thus, under the supposition that x is a person, we are forced to reject any predication about y entailing that y falls in some sense short of being a person, and so we are forced to accept that y is a person. By *reductio*, the truth of the relevant instance of (N) follows.

Again, this irrelevance can be made to emerge also from a different angle. We may assume that, under certain circumstances, I should try my best to

¹¹Presumably, the predicate 'treated in way w ' (as opposed to 'to be treated in way w ') is precise in at least the relevant respects so as to obey classical logic, and hence, plausibly, there possibly is a finite series of the kind suggested in the text where there is a last element treated in way w . But that just shows that we can be forced to discriminate against our own convictions, and it is absurd to seek any safeguard against this (sadly real) possibility in features of the use of a word (such as its vagueness). It doesn't show that these convictions are wrong. We can still insist that no discrimination should be made on the basis of such an invidious difference, even though, when forced to cope with all the elements of the series in the same situation, we are forced to operate such a discrimination. See chapter 6 for a thorough discussion of the theoretical aspects of this phenomenon.

save x 's life *because* x is a person. But, certainly, it is not the case that I should do so because x 's atoms are arranged in at most such and such distances from one another and it is not the case that the distance of even just two of them is a nanometre greater: in such a context, that x falls on one side rather than the other of such an exquisite distinction has no practical grounding interest,¹² as opposed to her falling on one side rather than the other of the distinction between being a person and falling in some sense short of being such. Thus, by *reductio*, the property of being a person cannot be coextensive with any sharply bounded property. Since it would be so coextensive if the relevant instance of (N) were not to be accepted, the relevant instance of (N) is to be accepted.

Feeling. Suppose that x is a Gothic cathedral. Suppose that y differs from x at most for the fact that one of y 's atoms is in a location which is within a nanometre distance from the location correlating to the location of the corresponding atom part of x . Then it should be correct to assert that y is a Gothic cathedral as well—*anything* falling short of this would seem to draw a *preposterous* difference between x 's and y 's artistic status based on a difference, such as the nanometrical displacement of a single atom, which we perceive to be fully *irrelevant* for Gothic cathedrals.

Again, the reason for the perceived irrelevance seems clear enough. Architectural styles of the relevant kind are individuated according to their embodying in a concrete form the spirit of an age. That something is a Gothic cathedral implies that it embodies in its forms the ideas and feelings of an age, which revive in the admiring eyes of the careful beholder, activating in her a particular emotional state. Such a state, or anything similar to it, would be grossly inappropriate for anything falling in some sense short of being a Gothic cathedral: only deceived or uneducated persons could feel something even remotely similar for something which falls in some sense short

¹²In this and in the following example, we shift our focus from explanation through causes to grounding through reasons, while remaining neutral on the relation between these two types of understanding.

of being a Gothic cathedral. But no nanometrical difference can warrant the two vastly different emotional states required, respectively, by the contemplation of a Gothic cathedral and the contemplation of something which, in effect, comes down to a failed Gothic cathedral—on the contrary, only a smooth modulation is permitted. Thus, under the supposition that x is a Gothic cathedral, we are forced to reject any predication about y entailing that y falls in some sense short of being a Gothic cathedral, and so we are forced to accept that y is a Gothic cathedral. By *reductio*, the truth of the relevant instance of (N) follows.

Again, this irrelevance can be made to emerge also from a different angle. We may assume that, under certain circumstances, I should find pleasure in walking around inside x *because* x is a Gothic cathedral. But, certainly, it is not the case that I should find pleasure in doing so because x 's atoms are arranged in at most such and such distances from one another and it is not the case that the distance of even just two of them is a nanometre greater: in such a context, that x falls on one side rather than the other of such an exquisite distinction has no emotional grounding interest, as opposed to its falling on one side rather than the other of the distinction between being a Gothic cathedral and falling in some sense short of being such. Thus, by *reductio*, the property of being a Gothic cathedral cannot be coextensive with any sharply bounded property. Since it would be so coextensive if the relevant instance of (N) were not to be accepted, the relevant instance of (N) is to be accepted.

2.2.2 Stretching the Truth

In this section, I aim at capitalizing on the phenomenon that we are sometimes willing to *stretch* the information that we gather about some cases to other cases that are similar but not necessarily identical to them in the relevant respects, and that we take such stretching to be *indefeasible*. If we are told that Bonn is very far from Berlin and know that Cologne is very close to

Bonn, we indefeasibly conclude that Cologne is also very far from Berlin; if we are told that 10 kilometres is a very long distance to run, we indefeasibly conclude that 9.999 kilometres is a very long distance to run; if we are told that 10 hours of work per day is too much, we indefeasibly conclude that 9 hours, 59 minutes and 59 seconds of work per day is too much. Here, that the stretching is indefeasible only means that it is *bound to be true if the original information is true*, and the phenomenon is that we usually take stretching to be indefeasible in this sense. Emphatically, the phenomenon is not that under no circumstances may the warrant a speaker has for the stretching be defeated by (very good) misleading evidence. For example, one can certainly be struck by a sorites paradox for ‘very close’ in such a way as to lose one’s warrant for inferring ‘Cologne is very far from Berlin’ from ‘Bonn is very far from Berlin’ and ‘Cologne is very close to Bonn’, just as one can certainly be struck by the Liar paradox in such a way as to lose one’s warrant for inferring ‘‘Snow is white’ is true’ from ‘Snow is white’.

Let us deepen our understanding of the phenomenon of stretching by focussing on a particular case. Suppose that I tell you that arriving at time t (specified on a second scale) is arriving roughly on time. Then it seems that you can indefeasibly conclude that also arriving at $t + 1$ is arriving roughly on time. More generally, given that the veridical information that arriving at t is arriving roughly on time has been gathered, the inference to the conclusion that also arriving at $t + 1$ is arriving roughly on time seems to be indefeasibly warranted. The claim is not just that, in such a situation, the *inference* cannot fail to be *truth preserving*,¹³ but that, in such a situation, the *conclusion* of the inference (namely, that arriving at $t + 1$ is arriving

¹³Throughout, an inference or an argument is understood to be *truth preserving* (in a situation) iff, (in that situation) *if* all the premises are true, so is the conclusion (that is, in terms of the truth (in that situation) of a suitable conditional). A decent underlying truth theory is presupposed, but truth preservation will not be assumed to be either necessary or sufficient for validity. The rather subtle distinction just drawn in the text does matter in this context, as I hope it will become clear shortly (see section 5.3.2 for more on truth preservation).

roughly on time) cannot fail to be *true*.

Interestingly, there seems to be a restriction on the way the initial information has to be gathered. Almost any way, whether non-inferential (*via* testimony, or memory, or perception, or introspection, or intuition) or inferential, will license the inference, *unless* it turns out ultimately to rely on an analogous inference from the information that arriving at $t - 1$ is arriving roughly on time. For suppose that I came to believe that arriving at t is arriving roughly on time just because (in the rational sense of ‘because’) someone else first veridically intuited and told me that arriving at $t - 1$ is arriving roughly on time and I subsequently inferred from that that also arriving at t is arriving roughly on time. In such a situation, although it seems that *my* inference cannot fail to reach a true conclusion, it does not seem that *your* inference is in turn guaranteed to reach a true conclusion—only the unstretched truth is allowed to be stretched.

This restriction invites the following puzzle. In the situation envisaged, it is true that arriving at t is arriving roughly on time (since the situation is such that both it is true that arriving at $t - 1$ is arriving roughly on time and it is true that, if arriving at $t - 1$ is arriving roughly on time, then arriving at t is arriving roughly on time), yet it might be untrue in it that arriving at $t + 1$ is arriving roughly on time. Under the very plausible assumption that *modus ponens* is unrestrictedly valid, is it then untrue in that situation that, if arriving at t is arriving roughly on time, then arriving at $t + 1$ is arriving roughly on time, since—so we have been taught—truth in a situation must be closed under logical consequence? That would be intolerable, as it would fail to verify the relevant instance of (N), and so would give rise to situations where it would not be correct to say that there is no difference between t and $t + 1$ as far as arriving roughly on time is concerned.

But it needn’t be so. As I will elaborate in section 5.4.6, truth in a situation may neutrally be thought to be determined by the logical consequences of an *initial* collection of truths specifying that situation. *If* the operative consequence relation is transitive, then truth in a situation is indeed closed

under logical consequence and the relevant instance of (N) will fail to be true in the situation envisaged (since, otherwise, by *modus ponens*, it would also be true that arriving at $t + 1$ is arriving roughly on time). But, as we will see in chapter 4, the weakenings of classical logic most friendly to (N) are such that the consequence relation is *not* transitive. Without transitivity, the truths that arriving at $t - 1$ is arriving roughly on time, that if arriving at $t - 1$ is arriving roughly on time, then arriving at t is arriving roughly on time and that, if arriving at t is arriving roughly on time, then arriving at $t + 1$ is arriving roughly on time simply do not entail that arriving at $t + 1$ is arriving roughly on time, and so do not require its truth.

They do require it jointly with the further assumption that arriving at t is arriving roughly on time, which is indeed true in the situation envisaged. But, as has already been remarked, the (rejected) requirement that truth in a situation be closed under logical consequence follows from the (accepted) requirement that the logical consequences of an initial collection of truths specifying a situation be themselves true in that situation only under the (rejected) requirement that logical consequence be transitive. Of course, in the situation envisaged, it might be the case that a conditional with a true antecedent and an untrue consequent is true, and that there is a last time t such that it is true that arriving at t is arriving roughly on time but untrue that arriving at $t + 1$ is arriving roughly on time—but that just goes with the territory once we are working in a classical metatheory of truth in a situation, and only shows that the connection between absolute semantic notions such as truth and untruth and the corresponding situation-relative notions is, at best, not immediate (again, see section 5.4.6 for elaboration of this point).

Indeed, the crucial epistemic difference drawn earlier between inferences that rely on (N) from premises not ultimately dependent on an analogous inference and inferences that rely on (N) from premises that are so ultimately dependent can be seen as the flip side of the crucial semantic difference just drawn between the stronger condition of being a logical consequence of an initial collection of truths specifying a situation and the weaker condition

of being a logical consequence of a collection of truths in that situation. In section 5.4, we will explore how both these differences can be best made sense of and grounded in a logical framework where the consequence relation is not unrestrictedly transitive.

We still do not have quite what we want, since the relevant instance of (N) (that, for every t , if arriving at t is arriving roughly on time, then arriving at $t + 1$ is arriving roughly on time) does not immediately follow from the fact that, for every t , if the veridical information that arriving at t is arriving roughly on time has been gathered, then arriving at $t + 1$ is arriving roughly on time.¹⁴ But suppose that arriving at t is arriving roughly on time. Then it is certainly possible, at least in most cases, that it is said to you (knowledgeably or not) that arriving at t is arriving roughly on time. Indeed, most veridical pieces of information can be acquired in one way or another (knowledgeably or not). The principle just established yields then that it is possibly the case that arriving at $t + 1$ is arriving roughly on time, and so, since the ‘possibly’-operator can be deleted in conceptual matters such as this, that it is the case that arriving at $t + 1$ is arriving roughly on time. Thus, under the supposition that arriving at t is arriving roughly on time, we are forced to accept that arriving at $t + 1$ is arriving roughly on time. By *reductio*, the truth of the relevant instance of (N) follows.

Two alternative, more conservative explanations of the phenomenon of stretching might seem tempting. Firstly, one might think that, in the relevant cases, knowledge (and other epistemic properties such as warrant) is governed by a *margin-for-error principle*, so that, for every t , knowledge that arriving at t is arriving roughly on time requires truth that arriving at $t + 1$ is arriving roughly on time (Williamson [1992]; Williamson [1994], pp. 216–47; Williamson [2000b], pp. 93–134 and Mott [1998]; Williamson [2000a]; Sorensen [2007]; Williamson [2007] for a critical discussion). Given this, one could explain the indefeasibility of the inference by observing that, if the information that arriving at t is arriving roughly on time is knowledgeably

¹⁴Thanks to Crispin Wright for pressing this worry.

(or warrantably) gathered, then it must be the case that arriving at $t + 1$ is arriving roughly on time. However, such an explanation would seem to be radically incomplete: as I have already suggested, the inference seems to be no less indefeasible when the true information is gathered unknowledgeably (or unwarrantably) outside of the putative margin for error—the speech ‘I don’t care whether the source really *knows* that arriving at t is arriving roughly on time: as long as it is *true* that arriving at t is arriving roughly on time, it is also true that arriving at $t + 1$ is arriving roughly on time’ is perfectly natural. What triggers the indefeasibility of the inference seems to be the plain truth of the information rather than its knowledgeability (or warrant).

Secondly, one might think that the phenomenon can be explained in terms of the high, though usually < 1 epistemic probability of truth preservation by these inferences (Sorensen [2001], pp. 57–67, who still subscribes to the weaker claim that there can be no warrant for believing any counterexample to a relevant instance of (N)). The phenomenon would then be revealed to be of the same kind as the inference from someone’s having a ticket of a fair lottery with, say, 1,000,000 participants to her losing the lottery: this inference too enjoys a very high epistemic probability of truth preservation which plausibly makes it legitimate in thought (if not in speech). However, such a model does not seem to fit all the aspects of the phenomenon of stretching. In particular, the defeasibility of the (highly probably truth preserving) inference from someone’s having a ticket of a fair lottery to her not winning the lottery seems to be completely lacking in the case of stretching inferences. In turn, this difference is reflected in the fact that the conditionalizations of the latter inferences support the corresponding universal closure (their epistemic probability of truth preservation is not higher than the epistemic probability of ‘For every time t , if arriving at t is arriving roughly on time, then arriving at $t + 1$ is arriving roughly on time’) whereas the conditionalizations of the former do not (their epistemic probability of truth preservation is higher than the epistemic probability of ‘Everyone with a ticket of a fair lottery will

lose'). The difference also reflects itself in the fact that the speech 'Arriving at t is arriving roughly on time. Therefore, arriving at $t + 1$ is arriving roughly on time' is acceptable whereas the speech 'Henry has a ticket of a fair lottery. Therefore, Henry will lose the lottery' is not.

Once the existence of stretching inferences has been established, it is instructive to investigate which purposes they serve. Though possibly very *similar*, most things are not *exactly* alike in any respect. Gathering information about a particular item, we should wish to be in a position to apply it to every item relevantly similar to it (even if not as good as the original item): our extrapolations about the world would be seriously hampered if exact similarity were required, for the new things encountered are very seldom exactly similar in some respect to the old ones. Hence, it seems that there is a strong natural pressure towards adopting predicates which satisfy the relevant instances of (N): for there certainly is a strong natural pressure for gathering information in such a way as to make the widest possible use of it in the presence of new, almost certainly not exactly similar cases (that is, we want to be able to stretch the information), and the only way to achieve this is to conceptualize the information with concepts obeying the relevant instances of (N).¹⁵

Thus, to return to our case, there is a typical complex cluster of pieces of information that one can gather about a certain time t : that arriving at t is not arriving very late, that t is a right time at which to arrive, that people would usually react in a certain way were one to arrive at t etc. We should wish to be in a position to apply this wealth of information to every time relevantly similar to t (even if slightly later than it), and this is only possible if the several pieces of information are collected together into a concept (like the concept of being a time such that arriving at it is arriving roughly on

¹⁵Of course, the *width-of-use* requirement has to strike a balance with a contrary, but equally pressing, *informativity* requirement: while we want the information to be applicable to *enough* many cases, we don't want it to be applicable to *too* many of them (or, worse still, to *all* of them!).

time) which satisfies the relevant instance of (N). This is of course not to deny that the collecting concept may not be the only one to satisfy the relevant instance of (N)—each of the several pieces of information will have to be expressed with concepts that also do so if stretching is to be coherent in the first place. This further requirement is nicely predicted by our general explanation, as it follows from the special case in which the cluster of pieces of information we should wish to stretch is composed by a single piece.

Another at least equally important source of the need for stretching comes most clearly into view by reflection on the point of so-called “*vaguefiers*” in natural languages (see Lewis [1970]; Lakoff [1973]; Zadeh [1975]; Kamp [1975] for discussions and theories of vaguefication). One thinks that one can make it to the date by t , and makes consequent arrangements to get there at t . But one soon realizes that unforeseen but likely slightly delaying circumstances may occur, or simply that one’s belief to be able to make it by t might be slightly overoptimistic, due to likely subtle miscalculations. A rough-and-ready calculation shows that δ is a reasonable margin for error. The question then arises as to why one does not rest content with the promise of meeting between t and $t + \delta$, rather than vaguefying and promising to meet at *about* t ?

One minor reason is that any particular bounded interval would seem completely arbitrary, triggering undesired conversational implicatures (‘Why did he choose exactly ‘at $t + \delta$ ’ rather than ‘at $t + \delta + 0.000001$ ’?’ would ask herself the heedful lover). However, use of predicates which satisfy the relevant instances of (N) is by itself neither necessary nor sufficient to dispose of the arbitrariness induced by predicates that do not do so. It is not necessary because arbitrariness can be lost by underspecification rather than vaguefication, by choosing an appropriately coarse-grained classification of times (so that one can promise to arrive at 8 pm, meaning by this to arrive at any second between the beginning and the end of the hour). It is not sufficient because arbitrariness still attaches to ‘about t ’ (‘Why did he choose exactly ‘at about t ’ rather than ‘at about $t + 0.000001$ ’?’ would ask

herself the heedful lover) as long as ‘ t ’ designates a sufficiently specific time (see Alston [1964], p. 85; Sorensen [1989]; Burns [1995] for discussions of the relation between vagueness and underspecificity).

The major reason for preferring vaguefication lies elsewhere and can be seen as follows. Two contrasting requirements on fixing a time for a date can be identified:

- (a) In fixing a time, one commits oneself to be there at that time. This circumstance pushes towards a generously late time—one tends to make one’s life as easy as possible.
- (b) However, of course, too late a time may put into jeopardy some if not all the purposes of the date. This other circumstance pushes towards a not too generously late time—life is never too easy.

As a matter of empirical fact about our ordinary circumstances, δ will therefore be bound not to take into account all the unforeseen but likely slightly delaying circumstances which may occur, or all the likely subtle miscalculations that led to the choice of t in the first place. But this puts any candidate for being δ which is good enough at steering a middle course between the two just noted contrasting requirements (a) and (b) under a terrible pressure: for the arising of *any one more* of the unforeseen but likely slightly delaying circumstances or the existence of *any one more* of the likely subtle miscalculations will suffice to engender a failure to comply with one’s commitment. Clearly, shift to the successor candidate will do little to alleviate this pressure, and even that little is likely to be offset by a lower score on the dimension of requirement (b). What is needed to remove the pressure is of course a specification of the time which already includes a provision for close enough times (those that would be needed should any one more of the unforeseen but likely slightly delaying circumstances arise, or should any one more of the likely subtle miscalculations have taken place)—what is needed is a specification of the time with a predicate satisfying the relevant instances of (N).

The kind of situation envisaged can be modelled in the following way. For simple enough cases, we can assume that the set X of all such likely-to-happen impediments is such that all of its members enjoy more or less the same (high) epistemic probability and are independent from one another. δ can then be seen as allowing for a certain finite number i of such events to happen. Since δ is a best candidate, requirement (a) entails that the epistemic probability that all the members $x_0, x_1, x_2 \dots x_{i-1}$ of any $Y \subseteq X$ whose cardinality is i unluckily happen together is not too high, while requirement (b) jointly with empirical facts about our ordinary circumstances entails that it will not be as low as one might ideally wish. In many cases, this will arguably constrain the value of the epistemic probability of all the members of any such Y unluckily happening together in the neighbourhood of .2 (at least according to my personal estimate of the pros and cons!). But then there will be a .2 probability that one can only make it *just* on time—a .2 probability that a situation will be realized in which the highly epistemically probable happening of *just one more* impeding event will tilt the balance from complying with one's commitment to failing to comply with one's commitment. Even if low, a value of .2 still represents an unreasonably high risk of being in a situation in which one is very likely to be subject to the sad mock of failing to comply with one's commitment because of the occurrence of a single, minute, in itself insignificant impediment.

The shift to a predicate satisfying the relevant instance of (N) avoids this, since the time interval selected will now be such that, if it allows for a certain finite number i of impeding events to happen, it also allows for $i + 1$ such events to happen. This does not of course guarantee that one will arrive on time, nor that the epistemic probability of this not happening is as low as one might ideally wish (in the model just sketched, it can e.g. be set to be only slightly lower than .2), but it does ensure that the epistemic probability that one will be in a situation in which one is very likely to be subject to a sad mock is 0.

Might one not, after choosing an appropriate δ , add an additional, suit-

ably extended buffer ε and promise to meet between t and $t + \delta + \varepsilon$, while still aiming at arriving between t and $t + \delta$? Given the suitable extension of ε , one would thus still keep the promise even if one more impediment occurs other than those allowed for by δ .¹⁶ This strategy only works if $t + \delta + \varepsilon$ also satisfies requirement (b) and at the same time is such that the epistemic probability that a maximal number of impeding events compatible with it happen (and so the epistemic probability that one will be in a situation in which one is very likely to be subject to the sad mock of failing to comply with one's commitment because of the occurrence of a single, minute, in itself insignificant impediment) is not unreasonably high (i.e. significantly lower than .2). As I have already noted, in many cases, as a matter of empirical fact about our ordinary circumstances, there is no reason to believe that these two conditions can be jointly satisfied for any choice of δ and ε .

The major reason for preferring vaguefication in this case has thus been traced to the need of insuring oneself from failing to comply with one's commitment due to the occurrence of a single, minute, in itself insignificant impediment. Such a need can be satisfied only if the commitment is expressed using a concept which allows for the relevant stretching inferences. Having worked out the reasons for vaguefication with respect to a very particular example, it is easy to see how these reasons can be generalized to a wide range of cases. For many F , two contrasting requirements on accepting that x falls under the concept of being F can be identified:

- (a') In accepting that x falls under the concept of being F , one commits oneself to x 's being good enough as to meet a sufficient condition for falling under the concept of being F . This circumstance pushes towards generously relaxed sufficient conditions for falling under the concept of being F —one tends to make one's life as easy as possible.
- (b') However, of course, too relaxed sufficient conditions may put into jeopardy some if not all the purposes of applying the concept of being F .

¹⁶Thanks to Crispin Wright for suggesting this strategy.

This other circumstance pushes towards not too generously relaxed sufficient conditions—life is never too easy.

As a matter of empirical fact about our ordinary circumstances, any sharply bounded property candidate for being picked out by the concept of being F will therefore be bound not to take into account all of x 's unforeseen but likely deviations on any dimension of variation relevant for its being F . But this puts any such candidate which is good enough at steering a middle course between the two just noted contrasting requirements (a') and (b') under a terrible pressure: for x 's *slightest* deviation on any dimension of variation relevant for its being F will suffice to engender a mistake of some kind or other in the acceptance that x is F . Clearly, shift to a slightly more generous sharply bounded property will do little to alleviate this pressure, and even that little is likely to be offset by a lower score on the dimension of requirement (b'). What is needed to remove the pressure is of course a specification of the sufficient conditions for falling under the concept of being F which already includes a provision for slightly weaker sufficient conditions—what is needed is a concept satisfying the relevant instances of (N).

I would like to close the discussion of vaguefication by proposing a new argument for instances of (N) involving *hedging* vaguefiers (not all vaguefiers have hedging effects: for example, 'extremely' vaguefies precise predicates such as 'acute' (as applied to angles), but, clearly, 'extremely acute' does not hedge 'acute!'). On the one hand, it seems very plausible that at least some hedging vaguefiers have a finite least upper bound of hedging power: '6 feet tall' can be hedged by 'roughly 6 feet tall', but the idea that the latter can in turn be hedged by 'roughly roughly 6 feet tall' looks rather dubious. We do not seem to have a conception of a height which, whilst good enough to count as being roughly roughly 6 feet tall, is not good enough to count as being roughly 6 feet tall. In an intuitive sense of the word, 'roughly 6 feet tall' "fuzzified" the area sharply demarcated by '6 feet tall': what else remains for 'roughly roughly 6 feet tall' to do? We do not seem to have a conception

of a non-trivial fuzzification of a fuzzy area. On the other hand, ‘roughly’ does seem to obliterate the existence of minute differences, so that, for every F , if x is F and y differs only minutely from x in the relevant respects, y is roughly F . However, the two premises:

- (i) If x is roughly roughly 6 feet tall, then x is roughly 6 feet tall;
- (ii) If x is roughly 6 feet tall and y is 1 inch shorter than x , then y is roughly roughly 6 feet tall

entail the conclusion:

- (iii) If x is roughly roughly 6 feet tall and y is 1 inch shorter than x , then y is roughly roughly 6 feet tall,

which in turn entails, by *reductio*, the truth of the relevant instance of (N).

2.2.3 Application by Casual Observation

In this section, I will work from another remark made by Wright [1975] (cf Wright [1976]). In considering the high plausibility of principles of the form of (N) for a noun like ‘heap’, Wright briefly remarked that “[h]eap’ is essentially a coarse predicate, whose application is a matter of rough and ready judgement [...] [i]t would for example be absurd to force the question of the execution of the command, ‘Pour out a heap of sand here’, to turn on a count of the grains [...] our conception of the conditions which justify calling something a heap of sand is such that the justice of the description will be unaffected by any change which cannot be detected by *casual* observation” (Wright [1975], p. 335).

As it stands, I think that, suggestive as it may be, this remark is in need of crucial supplementation. For what it only shows is the desirability of predicates whose application can be decided by casual observation *for some*

cases—that this property does not force satisfaction of the relevant instance of (N) can be seen by reflecting that it is also possessed by predicates like ‘within a 1.171979 metre distance’ which uncontroversially fail to satisfy the relevant instance of (N) (cf Sainsbury [1995], pp. 27–8; Weintraub [2004], pp. 237–8). The question then naturally arises whether there is any reason relating to applicability by casual observation which would lead us to use a predicate which possibly satisfies the relevant instance of (N) (like ‘close enough’) rather than one which does not (like ‘within a 1.171979 metre distance’) (some reasons not so relating have already been explored in sections 2.2.1, 2.2.2).

I divide the argument in favour of a positive answer to the foregoing question in two legs. Let us call ‘an *occasion*’ any situation with respect to which a predicate is applied and let us call ‘a *case*’ any object to which a predicate is applied. Then, the first leg of the argument reflects on the fact that we do demand of some predicate that, *on many occasions*, there be a guarantee that it is possible to decide its application by casual observation (under normal, non-deceptive circumstances)¹⁷ *for every case* (rather than just *for some cases*). For example, we do assume that, on many occasions, there is a guarantee that it is possible to decide by casual observation, for every contextually relevant object x , whether x is close enough or not.¹⁸

Consider for instance the injunction:

(I₀) Slow down just in case an animal is close enough to the racing track!

On many occasions, (I₀) does not strike us as far-fetched at all (indeed, it is very often issued!), unlike the injunction:

(I₁) Slow down just in case an animal is within a 1.171979 metre distance from the racing track!

¹⁷In the following, I will leave this qualification implicit.

¹⁸Throughout, I assume of course that there is no problem in seeing the object itself.

Arguably, (I₁) strikes us as far-fetched because being within a 1.171979 metre distance is not the kind of fact that is guaranteed to be knowable by one's cognitive capacities when one is zooming on the racing track. If that is so, then presumably (I₀) does not strike us as far-fetched because the following *weak observability* principle seems to hold for 'close enough':

(WO) On many occasions, positive and negative cases of closeness enough are guaranteed to be knowable by casual observation.

Note that 'within a 1.171979 metre distance' does not satisfy the appropriate analogue of (WO) because, even if *in fact* there are only easy positive and negative cases of being within a 1.171979 metre distance, there is no *guarantee* that this will be so—there is no guarantee that it will not be the case that an object is, say, within a 1.171980 metre distance but not within a 1.171979 metre distance. The existence of a guarantee for positive and negative cases of closeness enough thus implies that there are no objects and distances such that it is not the case that, on at least some occasion, it is knowable whether that object at that distance is close enough (henceforth, (WO) will be understood as carrying this implication). If there were a range of possible such cases not so knowable (the "strong borderline cases" of closeness enough), (I₀) should strike us just as far-fetched as (I₁). But the fact is that it doesn't.

Isn't it the case that we are not struck because we are in some sense *ignoring* the possibility of strong borderline cases? Until clear independent evidence has been presented for the postulation of a mechanism which should trigger the ignoring, such a suggestion cannot be adequately discussed. But it is hard to believe that there is some such mechanism only for the strong borderline cases of 'close enough' and not for the hard cases of 'within a 1.171979 metre distance', as the suggestion would require if it is to be compatible with the asymmetry of our reactions to the two injunctions. And a dilemma seems to be lurking for any such suggestion. A supplementation of (I₀) with the injunction of shooting only ("*only*", not "*if*") at those animals

which are within range Δ (where Δ happens to be included in the range allegedly occupied by the strong borderline cases) does not seem to be vacuous. If the postulated mechanism is subtle enough so as to accommodate for this apparent non-vacuity, the supplemented injunction can be taken as the decisive evidence for the claim that, on many occasions, positive and negative cases of closeness enough are thought to be guaranteed to be knowable by casual observation.

Of course, this is not to deny that some cases might be harder than others to decide—such is the case for virtually every predicate of a natural language! What our acquiescing reaction to (I₀) does show however is that, hard as they may be, such cases are also understood to be guaranteed to be knowable by the expert eye of a good racer. Suppose that there are lots of animals at clearly different distances within Δ . The surprising fact is that (I₀) still does not strike us as far-fetched at all.

The second leg of the argument reflects on the very plausible claim that the following *strong indiscriminability* principle holds for ‘close enough’:

- (SI) For every x, y at nanometrically different distances, it is not the case that, [on at least some occasion, we can know by casual observation that x is close enough and, on at least some occasion, we can know by casual observation that y is not close enough].

Putting the two legs of the argument together, we can conclude that only predicates satisfying the relevant instances of (N) can be applied by casual observation in the sense required by (WO). For suppose that x is not close enough and that y is only a nanometre closer than x is. Then, by (WO), on at least some occasion it can be known by casual observation that x is not close enough, and so, by (SI), on no occasion can it be known by casual observation that y is close enough. But then, by (WO), y is not close enough. By *reductio*, the truth of the relevant instance of (N) follows (strictly speaking, the argument has been given only for ‘*not* close enough’, but I

argue in Zardini [2007a] that scruples about double-negation elimination are unmotivated in these cases).

It is tempting to try to reach the same result by appealing, instead of (SI), to the following *weak indiscriminability* principle for ‘close enough’:

(WI) On no occasion can we know that two objects are at nanometrically different distances.

Emphatically, (WI) does not entail that we cannot, *on different occasions*, know, of two objects x , y at nanometrically different distances, that x is close enough and that y is not close enough, which is what is needed for the previous argument to go through. This is so because we may well fail to be in a position to pool together the two pieces of information that x is close enough and that y is not close enough in order to derive, given plausible principles connecting closeness enough and distance, a difference in distance between x and y . For instance, our warrant for believing that x is close enough may be defeated by reflecting on the fact that we do not believe that y , whose distance is indiscriminable from x ’s, is close enough.

Crucially, we cannot hope to rescue the previous argument by only relying on (WI) as our indiscriminability principle while appealing instead to a stronger observability principle, like the *strong observability* principle:

(SO) On every occasion, positive and negative cases of closeness enough are guaranteed to be knowable by casual observation.

We cannot do so because the extra strength of (SO) as opposed to (WO) is not supported by the previous considerations concerning the asymmetry of our reactions to (I₀) and (I₁)—the asymmetry has been observed to occur only on *many* occasions, not necessarily on *every* occasion. Moreover, the extra strength of (SO) is not only unmotivated, but is in itself questionable. Suppose that there is a suitable series of animals approaching the racing

track each of which is suitably close to its immediate neighbours along the direction orthogonal to the racing track and suitably far from them along the direction parallel to the racing track. The surprising fact is that (I_0) now does strike us as just far-fetched as (I_1) (chapter 6 will try to provide an explanation for this). The impossibility of running a $(SO)/(WI)$ argument instead of a $(WO)/(SI)$ argument should not however distress us too much, given that the uncontroversial *finitude* of our discriminatory powers very plausibly sustains not just (WI) , but (SI) (see Wright [1987], pp. 239–43; Fara [2001], pp. 916–20 for some discussion on what this finitude exactly entails).

The preceding example hints at the point of having predicates whose application is on many occasions decidable for every case by casual observation. Most tasks involve *qualified mandate* for an action: the action should be performed if (qualified “*mandate*”) but also only if (“*qualified*” mandate) a certain condition obtains. For some such tasks, any condition would do, as long as its obtaining is at least in principle ascertainable (in the sense that, for every contextually relevant case x , a subject would sooner or later arrive at the correct answer to the question whether x satisfies the condition): a diligent subject in charge of the task can sooner or later ascertain for every case whether it satisfies or not the condition and consequently act on the instructions received (such is the task assigned to a god of destroying every house with 1,975 bricks in it).

However, at least for *finite* beings, many tasks involve *time constraints*, and for some of them these are such that the subject in charge of the task can only afford to observe casually her environment (such is the task assigned to a child of finding three small dogs). These latter tasks would rightly strike us as far-fetched if formulated with predicates not satisfying an appropriate analogue of (WO) , for they then would require a subject to do something there is no guarantee she will be able to do. Predicates satisfying the relevant instances of (N) find a source of their usefulness exactly in this kind of situation, as satisfaction of the relevant instance of (N) by a predicate is a

consequence of the predicate's satisfying appropriate analogues of (WO) and (SI) (of course, appropriate analogues of (WO) and (SI) are also satisfied by precise predicates which are true of either everything or nothing in the contextually relevant domain, but most qualified mandates require distinctions which cannot be drawn by such predicates).

2.3 Experiences Requiring the Absence of Sharp Boundaries

2.3.1 Seamless Change

In this section, I want to focus on what is probably one of the most basic ways in which the absence of sharp boundaries presents itself to a sentient creature capable of conceptualizing her experience.¹⁹ Some changes from red to orange are *seamless*. That is, in such changes, the object's change from being red to being orange seems to be accomplished only throughout (what one would intuitively consider to be) *the whole temporal stretch* of the change, or, at most, only in *considerably large* subintervals thereof. In other words, it seems that the change cannot be *located* in any *considerably smaller* subinterval of the whole temporal stretch of the change (let alone at any *instant* included in it). Indeed, if this were not the case, the very phrase 'the whole temporal stretch of the change', meant to pick out a quite extended temporal interval, would be a dramatic misnomer, as the real change would ultimately consist in a sudden, instantaneous jump from red to orange, with, strictly speaking, no real change before and after that (that is, no real change which affects an object's being red or orange). But, taking a cooling bar of iron to be the object changing from red to orange and δ a sufficiently small subinterval,

¹⁹Surprisingly, the phenomenon has never been clearly isolated for analysis in its specificity. Thanks to Crispin Wright for directing my attention to it (see his Wright [2007b] for some discussion).

nothing less than the truth of every instance of ‘It is not the case that the iron is red at t but not red at $t + \delta$ ’ seems to be required to rule out an unwanted sudden jump. This leads in turn to the relevant instance of (N) (note that the change need not occur along a *temporal* dimension: it may for instance occur along a *spatial* one, for example when the iron bar changes seamlessly from being red at one end to being orange at the other end).

The foregoing assumes for simplicity that the iron’s change is from being red to being orange, but an analogous point can obviously be made for any alternative candidate to being orange for being the property acquired by the iron as soon as it loses the property of being red. Even if it is (rather implausibly) contested that there is a *first* such property which is so acquired (because the set of relevant properties is not well-ordered by the *x-is-acquired-earlier-than-y* relation),^{20,21} the point can simply be restated as concerning the change from possession of the property of being red to lack of this very same property (rather than possession of any other property).

Note that rejection of the law of excluded middle would indeed allow for rejection of every instance of ‘The iron is red at t but not red at $t + \delta$ ’. One rejects of each member t of a collection of times that the iron is either red at t or not red at t . Assuming very plausibly that rejection of a disjunction is sufficient for rejection of both disjuncts, this commits one to rejecting, of each member t of such a collection, that the iron is red at t and to rejecting, of each member t of such a collection, that the iron is not red at t . We can assume that these times are strictly later than the times of which one asserts that the iron is red at them and strictly earlier than the times of which

²⁰Throughout, I will occasionally use hyphenated open sentences to denote properties and relations.

²¹Note that the point of the contention is not simply that there is no first “interesting” property that the iron acquires. *That* is indeed very plausible, for, given the continuity of the dimension along which the change occurs (a certain line through a section of the colour spectrum), a non-well-founded chain of “interesting” properties is readily available (consider the chain of properties denoted by ‘ ξ is no less orange than τ ’, where τ denotes a point on the dimension of the change).

one asserts that the iron is not red at them, that the length of the stretch constituted by these times is at least as great as any admissible appropriately small enough δ and, ignoring higher-order vagueness, that these three kinds of times are exhaustive on the relevant domain. Assuming very plausibly that rejection of either conjunct is sufficient for rejection of a conjunction, this commits one to rejecting every instance of ‘The iron is red at t but not red at $t + \delta$ ’. Interestingly, the converse implication (from rejection of every instance of ‘The iron is red at t but not red at $t + \delta$ ’ to rejection of the law of excluded middle) does not hold, as is witnessed by standard supervaluationist approaches (see e.g. Fine [1975]). Indeed, the target of the following four points is the more general position consisting in rejection of every instance of ‘The iron is red at t but not red at $t + \delta$ ’ and non-acceptance of the relevant instance of (N).

Arguably, rejection of every instance of ‘The iron is red at t but not red at $t + \delta$ ’ is not sufficient fully to capture the intuition of seamless change. For the intuition in question is, to repeat, that the change is not located in any considerably smaller subinterval of the whole temporal stretch of the change, which requires, as it were, a “positive” lack of change at any such subinterval: of any such subinterval, it is not just rejected that the change occurs in it (in the same way in which it is rejected that a borderline case of red is red, which usually goes together with the symmetric rejection that it is not red)—it is positively asserted that it is not the case that the change occurs in it. But such a “positive” lack of change of the iron between t and $t + \delta$ with respect to the property of being red requires that it be not the case that the iron is red at t but not red at $t + \delta$.

That simple rejection of every instance of ‘The iron is red at t but not red at $t + \delta$ ’ is not sufficient fully to capture the intuition of seamless change can also be seen by reflecting on the fact that it would not be sufficient to rule out every unwanted sudden jump *of some sort*. For simple rejection that the iron is red at t but not red at $t + \delta$ is consistent with acceptance that the iron is red at t but borderline red at $t + \delta$, where ‘borderline’ can be understood

in such a way as to make ‘ φ but borderline, φ ’ inconsistent, and hence in such a way as to make a process in which a red object becomes borderline red enough of a *change*.

Moreover, on this scheme, the quantified claim ‘For every t , it is not the case that the iron is red at t but not red at $t + \delta$ ’ is inconsistent with other uncontroversial assumptions, and so should be rejected, which seems to leave very little room for manoeuvre for preserving the idea that the change does not occur instantaneously.

Finally, on any way of implementing this scheme I know of, for some instance of ‘The iron is red at t but not red at $t + \delta$ ’, rejection of it will require rejection of its negation as well (the reason being, roughly, that at least one of the conjunct will fail to satisfy the law of excluded middle or will be gappy), which amounts to rejecting what seems to be a necessary condition for the change not to occur instantaneously at t .²²

The main thrust of the foregoing considerations is that we think of some processes as crucially *taking time*—milk doesn’t go off in a nanosecond, Sampras didn’t become a great tennis player in a nanosecond, I haven’t learnt English in a nanosecond—and, surprisingly enough, nothing less than the truth of the relevant instances of (N) seems to be able to entitle us to this very natural conception. Arguably, the conception is meant to be expressed in some uses of the phrase ‘seamless change’. It is essential to stress that not every use of this phrase can be made sense of as simply saying that the object in question is changing *continuously* along a certain dimension, exactly because some such use is meant to convey the negation of sudden jumps in the

²²The reader will have realized that analogues of all these four points apply to the analogous scheme considered in section 2.2.1, which was there criticized only using an analogue of the second last point. I present the whole battery of points here as I think that they are even more intuitive when applied to the phenomenon of seamless change. Indeed, one might argue (though I won’t attempt to do this here) that this particular phenomenon is somehow a paradigm for our general conception of tolerance, just as a venerable tradition has it that the phenomenon of continuous change is somehow a paradigm for our general conception of continuity.

exemplification of a property, jumps which are not ruled out by a simple continuous change along a certain dimension. For example, an object could be continuously increasing in temperature: this does not rule out jumps in the exemplification of the relevant heat-related properties, such as the property of being at most 20°C hot. The simple *continuity* of the function which in this case takes a time to the most specific heat-related property exemplified at it does not by itself ensure any *seamlessness* in the sense under discussion.

The apparatus of exemplifying a property *to a certain* (possibly normalized) *degree* could naturally be thought to capture this stronger sense of ‘seamless change’. Unfortunately, the implementation of the strategy faces severe difficulties. For starters, not all the properties to which the strategy should be applied can be associated with a suitable linear ordering. For example, x can be very good at spatial intelligence but not at numerical one, and y be *vice versa*. Intuitively, neither are x and y equally intelligent, nor is x more intelligent than y , nor is y more intelligent than x , even though the degrees to which they exemplify intelligence would have to be either equal or one greater than the other.

Even for those properties which can be so associated, no satisfactory answer has ever been provided to the question concerning the relation between exemplifying a property *simpliciter* and exemplifying it to a certain degree. On the one hand, to say that an object exemplifies a property p *simpliciter* iff it exemplifies p at least to degree δ makes no sense of seamless change, since the change from possession of p to lack of p will then be as instantaneous as the change from exemplifying p to degree δ to not doing so. On the other hand, to say that an object exemplifies a property p *simpliciter* iff it exemplifies p at least to a high degree simply shifts the problem of explaining the seamlessness of the change from possession of p to lack of p to the problem of explaining the seamlessness of the change from possession of the newly introduced property of [exemplifying p at least to a high degree] to lack of this latter property. Nothing has been gained.

Worst of all, the very same notion of exemplifying a property p to a

certain degree has usually been explained in such a way as to be reduced to exemplifying *simpliciter* the property of having a certain value along a certain dimension associated with p . This reduction implies that the change from exemplifying p to a certain degree to exemplifying it to a different degree, even though continuous, is no more seamless than the change of value along the relevant dimension. Since the latter does not by itself ensure seamlessness (see the third last paragraph), it is hard to see how the former could do so.

Those who reject (N) may still hope to make sense of the idea that at least the *preparation* for a change—if not the change itself—takes time. However, it is rather unclear what this idea consists in and how it is supposed to relate to the phenomenon of seamless change. Of course, from the perspective of a rejection of (N), the preparation for a change, as a kind of process, cannot come more seamlessly into existence than the change itself can. Still, it seems that the preparation itself might seamlessly come into existence, and the postulation, required by the present strategy, of another preparation for the original preparation's coming into existence looks dubious.

Moreover, even though, once existing, the process in which the preparation consists is allowed to stretch through time, it must be kept in mind that the property which is eventually going to be lost in the change is still present throughout the preparation for the change and that all the events participating in the preparation also consist in sudden jumps. Such a “preparation” does no better in suggesting a seamless change than the uneven journey of an old-fashioned minute hand from 1.00 pm to 1.29 pm does in suggesting a seamless change from the minute hand's not indicating 1.30 pm to its doing so.

Finally, while the idea of a preparation for a change makes intuitive sense in the case e.g. of the change consisting in the destruction of Carthage (think of the process of destroying houses, burning ships, deporting people), such an idea seems inapplicable to at least some other cases of seamless change. To return to our original example, there does not seem to be any similar preparation in the iron's changing from red to orange—no set of events which

jointly constitute the change even though each of them is in itself insufficient for doing so.

The same conclusions about seamless change can be reached by approaching the phenomenon from a different perspective. It is uncontroversial that we do *not perceive*²³ the boundaries required by what (N) negates. Of course, it is not in general the case that one's not perceiving that *P* entails one's perceiving that it is not the case that *P*. For example, being away from home, I may not perceive that my dog is at home without thereby perceiving that he is not. Yet, a case can be made that, in the case of our perception of the boundaries required by what (N) negates, the situation is usually underdescribed by simply saying that we do not perceive sharp boundaries (I thus disagree with Wright [2007c], p. 25, who thinks that this is all that is warranted by the phenomenology). For, at least in some cases, for every pair of neighbouring objects, we do *perceive* that the boundary of the exemplification of a property does *not* fall between them.

For example, in a well-executed slow *sfumando* from *forte* to *piano*, we can perceive, of any two neighbouring enough moments, that the boundary between the orchestra playing *forte* and its not doing so does not fall between them. That one can perceive *this* over and above one's not perceiving that the boundary falls between them seems to be what warrants a favourable aesthetic judgement in the first place. For one would not usually think of one's evidence for such a judgement to be, peculiarly enough (as opposed to many other very similar aesthetic judgements, like the judgement that the winds are playing too loud), distinctively *second-order* about one's lack of certain perceptions, which is all the evidence that would be afforded were it the case that a sharp boundary is only not perceived to exist rather than also perceived not to exist. And one would not usually think that the orchestra played simply deftly enough to make it the case that one did not perceive a boundary where there might well have been one—rather, the aesthetic

²³Here, I use 'perceive' and its like in a non-factive sense, understanding it as synonymous with lengthier constructions such as 'it looks to one as though'.

judgement is issued only because one would think that the orchestra played deftly enough not to make at any time a sudden jump from its playing *forte* to its not doing so.

Similar considerations can also be advanced with regard to one's appreciation of perceptually manifest features of non-aesthetic objects. For example, upon leaving a fuzzy cloud on a flight, one can perceive, of any two neighbouring enough spatial regions, that the boundary between the cloudy region and the non-cloudy region does not fall between them. That one can perceive *this* over and above one's not perceiving that the boundary falls between them seems to be what warrants a judgement of fuzziness concerning the cloud's boundaries in the first place. For one would not usually think of one's evidence for such a judgement to be, peculiarly enough (as opposed to many other very similar judgements concerning the cloud's shape properties, like the judgement that it is big), distinctively *second-order* about one's lack of certain perceptions, which is all the evidence that would be afforded were it the case that a sharp boundary is only not perceived to exist rather than also perceived not to exist. And one would not usually think that the water droplets were simply increasing smoothly enough in their density to make it the case that one did not perceive a boundary where there might well have been one—rather, the judgement of fuzziness is issued only because one would think that the droplets were increasing smoothly enough in their density not to make at any point a sudden jump from the cloudy region to the non-cloudy region.

Of course, one can accept all these points about the correct description of our phenomenology and still reject (N), on the grounds that our experience systematically deceives us in this regard. On this view, our experience of seamless change would be similar to our experience of geometrically impossible situations (as nicely exemplified e.g. in the works of Escher): even though both real *qua* experiences, what they represent is something that cannot be the case. Such a move depends on the availability of independent grounds for thinking that what is represented is impossible (which do indeed exist in

the case of experiences of geometrically impossible situations). I doubt such grounds exist in the case of seamless change (see chapter 4 for discussion of the most plausible candidate, the sorites paradox). In any event, recall that the aim of this chapter is not so much that of establishing (N) beyond any reasonable doubt, but to expose the unpalatable commitments incurred by rejecting it—in this case, the unpalatable commitment to a new aspect of systematic illusion in the way we experience the world.

2.3.2 Appearances

In this final section, I wish to trace a fairly specific and unusual argumentative path through the unwieldy jungle of questions concerning phenomenal entities (some of the main works relevant to the issues I will touch on are Goodman [1951]; Armstrong [1968]; Jackson and Pinkerton [1973]; Dummett [1975a]; Wright [1975]; Wright [1987]; Peacocke [1981]; Linsky [1984]; Travis [1985]; Hardin [1988]; Williamson [1990]; Raffman [2000]; Fara [2001]). I would like to distinguish between the *phenomenal* identity of x and y with respect to a certain quality q (the fact that x and y *appear* to be the same specific way with respect to q) and the *epistemic* identity of x and y with respect to q (the fact that x and y *are not known not* to be the same specific way with respect to q) and only focus on the former (see Chisholm [1957] for a canonical statement of the distinction—usually associated with an ambiguity in the word ‘look’—and Breckenridge [2007] for critical discussion). I will first defend the claim that apparent (i.e. phenomenal) identity requires identity (of appearances). I will then contend that apparent identity is preserved across minute enough differences of the relevant quality. I will finally show how these two claims in turn entail the relevant instance of (N), or something close enough.

Consider the following general *abstraction* principle about appearances (see chapter 4 for a logical treatment of abstraction principles affected by vagueness):

(APP) For every subject s , time t , quality q and objects x, y , the appearances of x 's and y 's q for s at t are the same iff [x and y would appear to s at t to be the same specific way with respect to q if presented to s at t].

For example, as applied to shade appearances, (APP) yields:

(APP^{shade}) For every subject s , time t and objects x, y , the appearances of the shades of x and y for s at t are the same iff [x and y would appear to s at t to be of the same shade if presented to s at t].

Before proceeding with my argument, some remarks on the formulation of these abstraction principles are in order. Firstly, while the occurrence of ‘same’ on the left-hand side of (APP) and (APP^{shade}) denotes *numerical identity* (among certain exotic objects, appearances), the occurrence of ‘same’ on their right-hand side (in the scope of ‘appear’) denotes *qualitative identity* (among common-and-garden objects, percepts). I will remain neutral here as to whether qualitative identity can itself be reduced to numerical identity (e.g. between properties). Secondly, the appearances referred to on the left-hand side are relativized to whichever quality (shade, shape, sound etc.) is referred to on the right-hand side. Thirdly, the counterfactuality of the right-hand side is needed in order to be able to determine appearances to s at t also for objects which are not perceived by s at t . This is of course not to rule out that there might be some benign indeterminacy in some instances of the relevant counterfactual conditional, which would then entail some benign indeterminacy in the corresponding identity statements concerning appearances. Fourthly, hereafter I will mostly leave implicit the antecedent of the relevant counterfactual conditional (leaving only the modal ‘would’ to indicate the intended counterfactuality) and the relativization to a subject, a time and a quality. Finally, I will assume as unproblematic the left-to-right direction of the abstraction principles in question.

I now turn to the first step of the argument, arguing in favour of the right-to-left direction of (APP). This direction seems essential in grounding

the right identity conditions for appearances. For what *explains* the fact that two objects would appear to be the same specific way? It cannot be just the fact that their appearances are merely *similar*, because appearances are often similar without determining that the objects they are appearances of would appear to be the same specific way (for example, the appearance of the shape of a collection of 5 grains is very similar to, but not identical with, the appearance of the shape of a collection of 6 grains, but a collection of 5 grains would appear to be of a different shape than a collection of 6 grains). Moreover, even if mere similarity in appearances between two objects could sometimes *negatively* determine that it is not the case that one would appear to be not the same specific way as the other, it is hard to see how it could *positively* determine that one would appear to be the same specific way as the other.²⁴ It is hard to see, more generally, how any fact falling short of a (numerical) identity could determine that an object would appear to be the same specific way as another object.²⁵ But, certainly, appearances are the kind of objects related by such identities if anything is.

Indeed, not only does identity of appearances ground what would otherwise seem to be a (theoretically highly undesirable) primitive appearing of a (qualitative) identity, thus constituting the ultimate explanatory basis of the *explananda* with respect to which appearings of (qualitative) identities are usually appealed to—it is also required for some explanatory work which cannot be carried out by any appearing of a (qualitative) identity. For example, the dog may react in the same specific way upon numerically different,

²⁴Again, elaborating on a point already made in section 2.3.1, it is not in general true that its being the case that *x* would not appear to be not the same specific way as *y* entails its being the case that *x* would appear to be the same specific way as *y*—for example, it is not the case that an atom of hydrogen would appear to me now to be of a different shape than an atom of oxygen although it is also not the case that it would appear to me now to be of the same shape as an atom of oxygen.

²⁵I should stress that the principle appealed to in the text, while very attractive for appearings, is doubtlessly rather improbable for other mental states such as e.g. believings: it is not at all hard to see how some fact falling short of an identity could determine that an object would be *believed* to be the same specific way as another one.

temporally remote but qualitatively very similar calls of his master. There is no relevant connection between the two phenomenological states the dog undergoes, let alone a phenomenological state in which it appears to the dog that one object (the first call) sounds the same specific way as another (the second call)—that is, a phenomenological state which would constitute an appearing of a (qualitative) identity. Only the simple, *unapparent* identity of the appearances of the sounds of the two calls is there to explain the identity of the specific reactions of the dog (an analogous example can be given for the inter-subjective rather than inter-temporal case).²⁶ But once identity of appearances has to be admitted in order to explain such inter-temporal and inter-subjective cases, it would become arbitrary not to admit it in order to explain the case where, for some s and t , two objects would appear to s at t to be the same specific way if presented to s at t .

The next step of the argument exploits a minimal connection between the epistemic notion of justifiedly believing and the phenomenal notion of appearing, to reach the conclusion that, under normal circumstances, given only minute enough differences between x and y with respect to the relevant quality,²⁷ an apparent (qualitative) identity between x and y must hold—that is, for every subject s and time t constituting a normal circumstance, if x and y differ minutely enough, x and y would appear to s at t to be the same specific way if presented to s at t . I argue for this conclusion by first observing that it should be uncontroversial that, under normal circumstances, minutely enough differing objects presented pairwise would be *justifiedly believed* to be the same specific way, in the sense that the belief that they are the same specific way would be *positively supported* by the available phenomenal evidence. That is, it is not just that the phenomenal evidence is merely *consistent with* the objects' being the same specific way—

²⁶Of course, in such cases the identity of the relevant appearances is not entailed by anything like (APP), which only concerns identity of appearances to the *same* subject at the *same* time. I believe that this shows that, important as they are, abstraction principles like (APP) do not exhaust our understanding of appearances.

²⁷Relativization to the relevant quality will be left implicit in the following.

rather, the phenomenal evidence also *points to* the objects' being the same specific way (in the sense of *favouring* the hypothesis that the objects are the same specific way over its negation). This should be uncontroversial once "normal circumstances" are glossed as, roughly, those where the perceiver is in optimal conditions for observing each pair of neighbouring objects in the relevant series (such gloss gets around the difficulties raised by Fara [2001], pp. 916–20).

However, if so much is granted, to reject that, under normal circumstances, minutely enough differing objects presented pairwise would appear to be the same specific way would commit one to maintaining that, under normal circumstances, there could be a case of minutely enough differing objects presented pairwise where one is justifiedly believed, on the basis of the phenomenal evidence, to be the same specific way as the other despite its not appearing to be such! What could the source of this justification be, if it is to be phenomenal but still fall short of being the appearing of a (qualitative) identity? It would rather seem that nothing less than the appearing *of a (qualitative) identity* could be an appearing which favours the hypothesis that the objects are the same specific way over its negation, and thus justifies the belief that they are so. Hence, the uncontroversial fact that, under normal circumstances, if x and y differ minutely enough, the belief that x and y are the same specific way would be justified seems to imply that, under normal circumstances, if x and y differ minutely enough, x would appear to be the same specific way as y .

Now, drawing the previous two threads together, suppose for concreteness that, under normal circumstances, x and y differ minutely enough in shade. Then, as I have contended in the second step of the argument, x would appear to be of the same shade as y , and so, by the right-to-left direction of ($\text{APP}^{\text{shade}}$) I have defended in the first step of the argument, the appearance of the shade of x is the same as the appearance of the shade of y . Therefore, under normal circumstances, if x and y differ minutely enough in shade, the appearance of the shade of x is the same as the appearance of the shade of

y —that is, for every subject s and time t constituting a normal circumstance, if x and y differ minutely enough, the appearance of the shade of x for s at t is the same as the appearance of the shade of y for s at t . Moreover, given the current understanding of ‘normal circumstances’, in this conclusion ‘ x ’ and ‘ y ’ range over all the elements of the relevant series, which could go for example from a clear case of red to a clear case of orange (note that this argument nowhere appeals to closure of appearing under logical consequence, and so does not fall prey to the objections levelled by Williamson [1994], pp. 180–4).²⁸

Strictly speaking, this conclusion does not yet vindicate (N). It would do so if, assuming that the appearance of the shade of x is a , we were allowed to substitute ‘ a ’ for ‘the appearance of the shade of x ’ at the last step. This would yield the conclusion that, if x and y differ minutely enough in shade and the appearance of the shade of x is a , then the appearance of the shade of y is a , which is in effect (very plausibly) logically equivalent with the instance of (N) obtainable by substituting ‘ ξ_0 and ξ_1 differ minutely enough in shade’ for ‘ ξ_0 *R*s ξ_1 ’ and ‘The appearance of the shade of ξ is a ’ for ‘ ξ is F ’. Unfortunately, the crucial substitution which is needed is an instance of the rule of indiscernibility of identicals, which is invalid in many logics for vagueness and so cannot neutrally be appealed to here (it is valid though in the final logics proposed in chapter 4). Be that as it may, even if possibly falling short of entailing the relevant instance of (N), the conclusion is certainly in keeping with the spirit of the naive theory of vagueness, since it does establish that appearances do not discriminate between objects which

²⁸Since the converse of this conclusion (namely, that, under normal circumstances, if the appearance of the shade of x is the same as the appearance of the shade of y , then x and y differ minutely enough in shade (if they differ at all)) should be uncontroversial and since identity is arguably transitive, this entails, under some natural assumptions, that x -differs-minutely-enough-from- y is transitive. That seems plausible given the vagueness of this relation (it would not of course be plausible for a relation which specifies precisely the amount of difference tolerated). Chapter 4 will show how this much transitivity can crucially still fall short of underwriting paradox.

differ minutely enough.

2.4 Conclusion

The foregoing arguments lend a very high plausibility to a great many instances of (N). In particular, they do so by showing the high value of predicates conforming to (N) in our thought about, experience of and interaction with the world. According to the naive theory of vagueness, conformity to (N) is, roughly, what the nature of the vagueness of a predicate is, and so, according to the naive theory, the high value so achieved—namely, the possibility of such thoughts, experiences and interactions—is what the point of vague predicates is. We have been exploring in some detail some of the different grounds of this value—these are then what, according to the naive theory, the sources of vagueness are. These sources, we have seen, are rooted in fundamental facts about our cognition and agency in the world. The foregoing arguments thus lend a very high plausibility to the naive theory itself. Not only is it the theory which is arguably explanatorily most powerful (see chapter 1)—it is the only theory which gives depth to vagueness and does not reduce it in the end to a rather uninteresting nuisance deriving in some way or other from our failure to codify explicitly sharp boundaries for our predicates. Indeed, since it is very plausible that vagueness would be conducive to the high value in question if anything were, the foregoing arguments show that rejecting the naive theory amounts to rejecting the very possibility of the achievement of such a value. This is a very high—in my view, too high—cost incurred by the rejection of the naive theory. The only serious defeater for the naive theory I know of is the already mentioned sorites paradox. If this can satisfactorily be solved, the vindication of the theory will almost be complete. To this dialectic we turn in the next chapter.

Chapter 3

Higher-Order Sorites Paradox

3.1 Introduction and Overview

Appealing as it may appear at a first glance, the naive theory of vagueness has been on reflection rejected by almost every commentator on the grounds that it is subject to standard sorites paradoxes (see chapter 4 for a presentation and discussion of these). Opponents of the naive theory have thus sought to find a suitably weaker claim which, while no longer subject to sorites paradoxes, still manages to capture (a great deal of) what the naive theorist tries to capture with her claim of tolerance. Within the dominant approach, the natural fall-back has been a claim of borderlineness. That such a retreat is really safe has been assumed without much argument, save for briefly noting that there does not seem to be any easy way of generating sorites paradoxes out of the materials afforded by borderlineness claims. The alleged advantage of the dominant approach over the naive theory has wholly relied on this article of faith. It is high time to shake it.

The rest of the chapter is organized as follows. Section 2 recalls a surprising phenomenon of ignorance related to vagueness and sketches what account is given of it by the dominant approach. Section 3 rehearses another phenomenon of vagueness, higher-order vagueness, articulating and

defending two main claims about it. Section 4 uses these claims to develop a higher-order sorites paradox, showing that the dominant approach is after all no less paradoxical than the naive theory. Section 5 draws the conclusions which follow from the higher-order sorites paradox for the dialectic between the dominant approach and the naive theory.

3.2 Ignorance

3.2.1 Ignorance and Borderlineness

Consider the series S of natural numbers from 0 to 1,000,000, the predicate ‘A person with ξ hairs on her scalp is bald’ (henceforth ‘ $B\xi$ ’) and a conversational context (a fairly common one, I suppose) where 0 and 1,000,000 are, respectively, indisputable positive and negative cases for the application of ‘ B ’. If there is a boundary in S between the B s and the $\neg B$ s,¹ where does it lie? As noted in section 1.3, we simply seem to be unable to provide a knowledgeable identification of such a boundary: ‘ B ’ is vague.

It is generally agreed that our surprising inability to provide a knowledgeable identification of the boundary between the B s and the $\neg B$ s is due to the vagueness of ‘ B ’ (as stressed in section 1.3, this phenomenon of ignorance extends from lack of knowledge to a more general lack of warrant, but, in order to avoid complexities which seem irrelevant for the purposes of this chapter, we will focus only on this peculiar lack of knowledge and take it as our guide to vagueness). But how exactly is the latter supposed to explain the former? The dominant approach has it that the vagueness of ‘ B ’ explains our epistemic inability insofar as ‘ B ’ presents borderline cases² of application in S : objects which are neither definitely B nor definitely $\neg B$

¹To preserve clarity, in this chapter I will help myself to a moderate, hopefully self-explanatory, regimentation of my language into Loglish.

²Henceforth, ‘borderline case’ and its like is used as a short for ‘strong borderline case’ and its like.

(I will have something more to say on the relation between the property of being borderline and the property of being definite in section 3.3.1). Letting henceforth the quantifiers range over the elements of S and ‘ \mathcal{D} ’ be a ‘definitely’-operator, the *borderlineness* principle:

$$(B) \exists x(\neg \mathcal{D}Bx \wedge \neg \mathcal{D}\neg Bx)$$

is accepted.

Let me stress right at the outset that all the arguments to follow will be fairly neutral with respect to the specific interpretation of the ‘definitely’-operator (for the various options, see section 1.3), turning only on basic common features required by its use in describing borderline cases and on a minimal normal modal logic **KT** validating its *factivity* and *closure under logical consequence*:

$$(F) \vdash \mathcal{D}\varphi \supset \varphi;^3$$

$$(C) \text{ If } \varphi_0, \varphi_1, \varphi_2 \dots \vdash \psi, \text{ then } \mathcal{D}\varphi_0, \mathcal{D}\varphi_1, \mathcal{D}\varphi_2 \dots \vdash \mathcal{D}\psi.^4$$

But, again, how exactly is (B) supposed to explain our surprising epistemic inability? It is—to my knowledge—universally and—to my mind—very plausibly accepted that, if there is indeed a boundary between the B s and the $\neg B$ s, this has to lie in the borderline area of B ness, and that, if two objects are neither $\mathcal{D}B$ nor $\mathcal{D}\neg B$, it is not known that one is B and the other

³Throughout, I will use ‘ \vdash ’ to denote the contextually relevant consequence relation.

⁴(F) and (C) have both come into question in another area where the notion of being definite is supposed by many to do some helpful work—namely, in the area of the semantic paradoxes. On pain of revenge, classical theories of truth such as the one developed in McGee [1991] must reject the definitization of (F) and non-classical theories of truth such as the one developed in Field [2002]; Field [2003a]; Field [2003c]; Field [2007a] must reject (C). No matter what one thinks of these moves in the context of the semantic paradoxes, I don’t think that a similar rejection of either (F) or (C) can be motivated in the case of vagueness.

$\neg B$. (A stronger, almost universally accepted ignorance principle is that its being borderline whether something is B implies ignorance as to whether it is B . This principle has however been forcefully rejected by some theorists of vagueness (in addition to the works of Crispin Wright cited in section 1.3, see Dorr [2003]; Barnett [2007a]; Barnett [2007b]). Nevertheless, these theorists would still accept the weaker ignorance principle just mentioned.) Factivity of knowledge reduces then to inconsistency the set consisting in these two claims and in the further claim that it is known where the boundary between the B s and the $\neg B$ s lies. So far so good.

3.2.2 Indefiniteness and Tolerance

Now, reflect that S is such that the *trichotomy* principle:

$$(T) \quad \forall x \forall y (x < y \vee y < x \vee x = y)$$

obviously holds. Moreover, the meanings of ‘ B ’ and ‘ \mathcal{D} ’ have been sufficiently explicated in order to validate the *monotonicity* principle:

$$(M) \quad \forall x ((\mathcal{D}Bx' \supset \mathcal{D}Bx) \wedge (\mathcal{D}\neg Bx \supset \mathcal{D}\neg Bx')),$$

where ‘ \supset ’ is the standard successor functor, which is well-defined on S (for simplicity’s sake, I am acquiescing in the usual, harmlessly false assumption that baldness is just a matter of the number of hairs on one’s scalp, see section 2.1).

Theorem 3.2.1. *Together with (T) and (M), (B) entails the indefiniteness principle:*

$$(I) \quad \neg \exists x (\mathcal{D}Bx \wedge \mathcal{D}\neg Bx').$$

Proof. Suppose for *reductio* that there is a definite sharp boundary between the B s and the $\neg B$ s (that is, that there is an object a such that a belongs to

S and $\mathcal{D}Ba \wedge \mathcal{D}\neg Ba'$). Consider then an arbitrarily chosen object b belonging to S . By (T), b is either less than a or greater than a or identical with a . If b is smaller than a , then, by (M), b is $\mathcal{D}B$, and therefore $\mathcal{D}B \vee \mathcal{D}\neg B$. If b is greater than a , then, by (M), b is $\mathcal{D}\neg B$, and therefore $\mathcal{D}B \vee \mathcal{D}\neg B$. If b is identical with a , then, by Leibniz's Law, b is $\mathcal{D}B$, and therefore $\mathcal{D}B \vee \mathcal{D}\neg B$. (Note that intersubstitutability of identicals in ' \mathcal{D} '-contexts may very well result problematic, at least under certain interpretations of ' \mathcal{D} '. This delicate issue lies however outside the scope of this essay.) Therefore, reasoning by cases, b is $\mathcal{D}B \vee \mathcal{D}\neg B$. Therefore, by an uncontroversial De Morgan's Law, b is $\neg(\neg\mathcal{D}B \wedge \neg\mathcal{D}\neg B)$. Therefore, by universal generalization, everything is $\neg(\neg\mathcal{D}B \wedge \neg\mathcal{D}\neg B)$. Therefore, by an uncontroversial $\forall\xi\neg\varphi \Rightarrow \neg\exists\xi\varphi$ quantifier manipulation, nothing is $\neg\mathcal{D}B \wedge \neg\mathcal{D}\neg B$. Therefore, by existential instantiation, nothing is $\neg\mathcal{D}B \wedge \neg\mathcal{D}\neg B$. Contradiction with (B). Therefore, by *reductio*, there is no definite sharp boundary between the B s and the $\neg B$ s (cf Greenough [2003], pp. 270–1).

□

(I) is certainly not as straightforwardly paradoxical as the stronger, naive *tolerance* principle:

(TOL) $\neg\exists x(Bx \wedge \neg Bx')$.

On the one hand, (TOL) says that at no point in S does a single step (that is, a step from a number to its successor) bring us from the B s to the $\neg B$ s, which—by the standard sorites paradox—seems to be inconsistent with 0's being B and 1,000,000's being $\neg B$. On the other hand, (I) only says that at no point in S does a single step bring us from the $\mathcal{D}B$ s to the $\mathcal{D}\neg B$ s, which seems to be consistent with 0's being B and 1,000,000's being $\neg B$, and may be thought adequately to describe what the apparent *smoothness* of S with respect to B ness consists in. However, some very plausible claims concerning the phenomenon of higher-order vagueness suffice to show that principles of the form of (I) (and, therefore, principles of the form of (B)) are, in a subtler

way, just as paradoxical as (TOL)—which, as remarked in section 1.4, was the most obvious candidate for a characterization of the nature of vagueness from which an explanation of our epistemic inability would nicely follow.

3.3 Higher-Order Vagueness

3.3.1 Being Definite^ω

In accounting for the effects in S of the vagueness of ‘ B ’, the new compound predicate ‘ $\mathcal{D}B$ ’ has been used by the dominant approach. As remarked in section 1.3, this predicate is itself vague: ‘ B ’ is higher-order vague. Two claims concerning higher-order vagueness seem to be highly plausible. Before stating and defending them, we need however to introduce some notation. For every natural number i , let ‘ \mathcal{D}^i ’ be a shorthand for the expression obtained by concatenating the empty string with i occurrences of ‘ \mathcal{D} ’. Assuming the availability in the object language of the expressive resources afforded by substitutional quantification, we can now introduce a ‘*definitely*^ω’-operator ‘ \mathcal{D}^ω ’ by letting ‘ $\mathcal{D}^\omega\varphi$ ’ be satisfied iff ‘ $\Pi i\mathcal{D}^i\varphi$ ’ is.

This is a particularly neutral way of introducing transfinite levels of the property of being definite without presupposing much with respect to the semantics of ‘ \mathcal{D} ’. On some semantics for a vague language (for example, supervaluationism and epistemicism), ‘ \mathcal{D} ’ usually receives an interpretation in terms of possible-world semantics, being in effect treated as a necessity-like operator. In such a semantic framework, ‘ \mathcal{D}^ω ’ can be interpreted as the *ancestral* of ‘ \mathcal{D} ’, thereby meaning that the possible-world semantics for ‘ \mathcal{D}^ω ’ uses as accessibility relation the ancestral (that is, the finite transitive closure) of the accessibility relation used by the possible-world semantics for ‘ \mathcal{D} ’ (Williamson [1994], p. 160). On other semantics for a vague language (for example, many-valued), ‘ \mathcal{D} ’ usually receives an interpretation in terms of algebraic semantics, being in effect treated as a lowering operation on the structure of values. In such a semantic framework, ‘ \mathcal{D}^ω ’ can be interpreted as

the *greatest lower bound* of ‘ \mathcal{D} ’, thereby meaning that the algebraic semantics for ‘ \mathcal{D}^ω ’ assigns to it the operation that takes a value v to the greatest lower bound of the set of values to which v is taken by the operations denoted by finite concatenations of ‘ \mathcal{D} ’ (Field [2007a], p. 34).

The first claim concerning higher-order vagueness is then that 0 (by assumption, an indisputable positive case of B_{ness}) is $\mathcal{D}^\omega B$ and that 1,000,000 (by assumption, an indisputable negative case of B_{ness}) is $\mathcal{D}^\omega \neg B$:

(O) $\mathcal{D}^\omega B0 \wedge \mathcal{D}^\omega \neg B1,000,000$.

The claim is validated by some of the most influential semantics for a vague language including a ‘definitely’-operator: supervaluationist semantics assigning super-truth to ‘ $\mathcal{D}\varphi$ ’ if super-truth is assigned to φ and allowing for ‘ $B0$ ’ and ‘ $\neg B1,000,000$ ’ to be super-true (see Fine [1975]), many-valued semantics assigning full truth to ‘ $\mathcal{D}\varphi$ ’ if full truth is assigned to φ and allowing for ‘ $B0$ ’ and ‘ $\neg B1,000,000$ ’ to be fully true (see Sanford [1975])—as well as, I suspect, being taken for granted in most of the literature (a notable exception is Williamson [1994], pp. 229, 232–3; Williamson [2002], pp. 145–6; Cian Dorr too has work-in-progress relevant to this issue). More importantly, the claim is very intuitive independently of the specific interpretation received by the ‘definitely’-operator (see section 1.3), for it amounts to saying that an indisputable positive or negative case does *not* exhibit vagueness at *any* order. Even more importantly, the claim seems unavoidable for a vague predicate like ‘ ξ is close to 0 and [ξ is far from 1,000,000 or identical with 0]’ as applied to 0 (positively) and 1,000,000 (negatively): this is so because it seems unavoidable that x -is-close-to- y is definitely $^\omega$ reflexive, that x -is-identical-with- y is definitely $^\omega$ reflexive and vacuous⁵ and that x -is-far-from- y is definitely $^\omega$ irreflexive. Finally, the claim can be justified by various considerations concerning the *function* which the ‘definitely’-operator is supposed to serve and which allegedly constitutes our primary grasp of it.

⁵A relation R is *vacuous* on a set X iff, for every $x, y \in X$, if Rxy , then $x = y$.

The issues here are surprisingly deep and difficult, deserving a far more extended treatment than I can afford in this essay, but let me sketch just one such consideration. Recall that, according to the dominant approach, borderline cases are supposed to be not just a theoretical construction needed, for every i , in the explanation of our epistemic inability concerning the boundary between the $\mathcal{D}^i B$ s and the $\neg \mathcal{D}^i B$ s; rather, they are supposed to be *manifested* in our experience of being confronted with “hard cases” of $\mathcal{D}^i B$ ness—that is, cases where, even after taking in all the relevant ‘ $\mathcal{D}^i B$ ’-free information, we competent speakers for ‘ $\mathcal{D}^i B$ ’ feel unconfident both in unqualifiedly applying ‘ $\mathcal{D}^i B$ ’ to the case in question and in unqualifiedly applying ‘ $\neg \mathcal{D}^i B$ ’ to it. Indeed, from within the dominant approach, it is arguable that, in general, our grasp of the ‘definitely’-operator flows from our grasp of a more basic ‘borderline’-operator, so that definite cases should be defined in terms of borderline cases by saying that an object is definitely F iff it is F and not borderline F (rather than, as usual, defining borderline cases in terms of definite cases by saying that an object is borderline F iff it is neither definitely F nor definitely $\neg F$).

By assumption, 0 is not such a case for ‘ B ’. So, go along S starting from 0 and considering, case by case, the application of ‘ B ’. Sooner or later, you will find yourself confronted with cases where, even after taking in all the relevant ‘ B ’-free information, you competent speaker for ‘ B ’ feel unconfident both in unqualifiedly applying ‘ B ’ to the case in question and in unqualifiedly applying ‘ $\neg B$ ’ to it: these are the borderline cases of B ness—that is, cases which are neither $\mathcal{D}B$ nor $\mathcal{D}\neg B$. By assumption, your epistemic situation with respect to 0 was *not like that*: reflecting on the *change* occurred between 0 and the borderline cases of B ness, you infer that 0 is *not* a borderline case of B ness, and therefore, by (F), that it is $\mathcal{D}B$.

More explicitly: reflecting on the change occurred between 0 and the borderline cases of B ness, you infer that 0 is not a borderline case of B ness—that is, 0 is $\neg(\neg \mathcal{D}B \wedge \neg \mathcal{D}\neg B)$. But 0 is B and therefore, by (F), it is $\neg \mathcal{D}\neg B$. Therefore, by adjunction and *reductio*, it is $\neg \neg \mathcal{D}B$. Therefore, by double-

negation elimination, it is \mathcal{DB} . Such a reasoning is of course intuitionistically invalid at its very last step. I leave it as an exercise to the reader to adapt the argument from here on in order to accommodate for this. However, note that, adopting instead the less usual definition of definite cases in terms of borderline cases suggested earlier, a legitimate worry concerning double-negation elimination is no longer possible, since, now, for something to be definitely F it takes no more than for it to be F and not to be a borderline case of F ness. (In general, no legitimate worry concerning double-negation elimination with respect to ' F ' is possible if ' F ' is known to be *exclusive and exhaustive* with respect to ' G ' over a particular range of application. By exclusivity, if something is G , it is $\neg F$, wherefore, by contraposition, if something is $\neg\neg F$, it is $\neg G$. By exhaustivity, if something is $\neg G$, it is F . By transitivity of the conditional, if something is $\neg\neg F$, it is F . See Zardini [2007a] for another application of this point in the case of vagueness.)⁶

You felt confident in asserting, on reflection, the difference between your epistemic situation with respect to 0 and your epistemic situation with respect to the borderline cases of B ness; since that there is such a difference is the only substantial assumption on which the previous conclusion that 0 is \mathcal{DB} depends, you (should) feel confident in asserting this conclusion. Hence, you (should) feel confident in applying ' \mathcal{DB} ' to 0. Indeed, it is just as clear that there is a difference between your epistemic situation with respect to 0 and your epistemic situation with respect to the borderline cases of B ness as is clear that 0 is B , and so your confidence in applying ' \mathcal{DB} ' to 0 should equal your confidence in applying ' B ' to it.

⁶The particular use of *reductio* on a premise got by adjunction is also relevantly invalid. Additional, plausible assumptions concerning intensional connections between 0's being $\neg\mathcal{D}\neg B$ and 0's being $\neg\mathcal{DB}$ would have to be made in order to turn the argument into a relevantly valid one. Alternatively, the same strategy of adopting the less usual definition of definite cases in terms of borderline cases would do in this case as well. I leave to the interested reader the details of this as well as those of similar modifications that would be needed at later stages of the argument. Thanks to Stephen Read for discussion of this point.

So, go along S starting again from 0 but, this time, considering, case by case, the application of ‘ DB ’. Sooner or later, you will find yourself confronted with cases where, even after taking in all the relevant ‘ DB ’-free information, you competent speaker for ‘ DB ’ feel unconfident both in unqualifiedly applying ‘ DB ’ to the case in question and in unqualifiedly applying ‘ $\neg DB$ ’ to it: these are the borderline cases of DB ness—that is, cases which are neither DDB nor $D\neg DB$. Having followed through the previous reasoning, your epistemic situation with respect to 0 was *not like that*: reflecting on the *change* occurred between 0 and the borderline cases of DB ness, you infer that 0 is *not* a borderline case of DB ness, and therefore, by (F), that it is DDB . (Comments analogous to those made earlier apply of course to this inference.)

You felt confident in asserting, on reflection, the difference between your epistemic situation with respect to 0 and your epistemic situation with respect to the borderline cases of DB ness; since that there is such a difference is the only substantial assumption on which the previous conclusion that 0 is DDB depends, you (should) feel confident in asserting this conclusion. Hence, you (should) feel confident in applying ‘ DDB ’ to 0. Indeed, it is just as clear that there is a difference between your epistemic situation with respect to 0 and your epistemic situation with respect to the borderline cases of DB ness as is clear that 0 is DB , and so your confidence in applying ‘ DDB ’ to 0 should equal your confidence in applying ‘ DB ’ to it. *Et sic in infinitum*. This “back-and-forth” argument establishes that, for every i , one should be confident in applying ‘ $D^i B$ ’ to 0—indeed, it establishes that one should be equally confident in applying ‘ $D^i B$ ’ to 0 as one is confident in applying ‘ B ’ to it. A parallel back-and-forth argument would of course establish the corresponding claim for ‘ $D^i \neg B$ ’ and 1,000,000. Together, these arguments commit one to (O).

3.3.2 Radical Higher-Order Vagueness

The second claim concerning higher-order vagueness is that, for every i , ‘ $\mathcal{D}^i B$ ’ is *definitely^ω vague*. We can establish this claim by reflecting that, for every i , we seem to be just as much unable to provide a knowledgeable identification of the boundary between the $\mathcal{D}^i B$ s and the $\neg \mathcal{D}^i B$ s as we are concerning the boundary between the B s and the $\neg B$ s (and *definitely^ω so*), and that there seems to be no relevant difference as regards the *source* of these various epistemic inabilities (and *definitely^ω so*). In particular, for every large i , it does not seem that the source of our epistemic inability with respect to the boundary between the $\mathcal{D}^i B$ s and the $\neg \mathcal{D}^i B$ s is constituted by our *computational* limits in understanding ‘ $\mathcal{D}^i B$ ’: rather, it seems that even suitable finite extensions of us would be in the same epistemic situation as we are with respect to vague expressions we do understand. Since the epistemic inability concerning the boundary between the B s and the $\neg B$ s is due to the vagueness of ‘ B ’ (and *definitely^ω so*), the epistemic inability concerning the boundary between the $\mathcal{D}^i B$ s and the $\neg \mathcal{D}^i B$ s is due to the vagueness of ‘ $\mathcal{D}^i B$ ’ (and *definitely^ω so*, by closure of the property of being definite—and, therefore, of the property of being definite^ω—under logical consequence). In other words, the vagueness of ‘ B ’ is *radical*:

(R) $\Pi i \mathcal{D}^\omega VAGUE(\mathcal{D}^i B)$.

Such a claim might however seem to be in contrast with at least one of the considerations adduced in support of (O), namely that regarding the possibility of predicates like ‘ ξ is close to 0 and [ξ is far from 1,000,000 or identical with 0]’ which force some cases (like 0) to be *definitely^ω positive* and some other cases (like 1,000,000) to be *definitely^ω negative*. The worry is that the straightforward argument which ensures in this and similar examples that there are *definitely^ω positive* and *negative* cases will also establish such cases as definite sharp boundaries at some higher order or other. Thus, in the case of ‘ ξ is close to 0 and [ξ is far from 1,000,000 or identical with 0]’,

it might be natural to think that 0 and 1,000,000 will be, for some i , the definite sharp boundaries between the $\mathcal{D}^i B$ s and the $\neg \mathcal{D}^i B$ s and between the $\mathcal{D}^i \neg B$ s and the $\neg \mathcal{D}^i \neg B$ s respectively (it will be even more natural to think so for those who are attracted to the mildly nihilist view that nothing is a definitely^ω positive or negative case of a normal vague predicate).

Focus on the ignorance phenomenon allows us a conclusive dismissal of the worry. For every i , the proponent of the worry grants that 0 is $\mathcal{D}^i B$. Were vagueness to peter out at the i th order in the manner envisaged, the proponent should be able knowledgeably to identify 0 as the sharp boundary between the $\mathcal{D}^i B$ s and the $\neg \mathcal{D}^i B$ s. Unfortunately for her, the fact is that she isn't. Each claim she might make to this effect will sound wholly preposterous—for all she knows, the sharp boundary will stabilize at 0 only at some order higher than the i th one.⁷ The worry also seems to miss completely the point of vague predicates associated with extremes on a certain dimension of comparison. At every order of the property of being definite, such a point is exactly that of singling out for special treatment not just the extreme, but an extended segment of the dimension (including the extreme). For every i , if ' $\mathcal{D}^i B$ ' were to apply positively only to 0, $\mathcal{D}^i B$ ness would completely pervert the point of B ness, which is to look at men with *very few* hairs on their scalp, not just at men with *no* hairs on their scalp (even though he has fewer hairs on his scalp, Yul Brynner is no more bald than Michail Gorbačëv is).

3.4 The Paradox

3.4.1 The Basic Form of the Paradox

We are now in a position to show that, because of the phenomenon of higher-order vagueness as described by (O) and (R), principles of the form of (I) are

⁷In conversation, Patrick Greenough has usefully dubbed this dimension of higher-order vagueness '*vertical higher-order vagueness*'.

just as paradoxical as (TOL). The paradoxicality of principles of the form of (I) in the presence of higher-order vagueness (*even without* (O) *and* (R)) has first been argued for by Wright [1987], pp. 262–6; Wright [1991], pp. 141–4, 150–60; Wright [1992], pp. 130–7. However, Wright’s argument needs a *prima facie* implausibly strong logic for the ‘definitely’-operator and is criticized by Edgington [1993], pp. 193–6; Heck [1993] on these grounds. The argument to follow, on the contrary, only needs the operator to be factive and closed under logical consequence (and hence targets also theories, such as the one offered in Cobreros [2007], which weaken the logic of ‘definitely’ in order to cope with the arguments just mentioned). The same paradoxicality has more recently been argued for by Fara [2003], pp. 196–205, whose argument, however, still needs a logic for the ‘definitely’-operator acceptable only on some of its specific interpretations (in particular, a logic according to which $\varphi \vdash \mathcal{D}\varphi$ is valid).

The paradoxicality of *margin-for-error* principles of the form of:

$$(ME) \quad \neg \exists x (\mathcal{D}Bx \wedge \neg Bx')$$

in the presence of higher-order vagueness (as described by (O) and (R)) has been argued for by Gómez-Torrente [1997], pp. 243–5; Gómez-Torrente [2002b], pp. 114–7, 119–24; Fara [2002] (to whom Williamson replies in his Williamson [1997], pp. 261–3; Williamson [2002], pp. 144–9). However, margin-for-error principles and their underlying epistemology are very controversial; by (F), they *entail but are not entailed by* principles of the form of (I), which, on the contrary, seem to be just platitudes about vagueness.

Our higher-order sorites paradox can most easily be made to emerge by using an appropriate, hopefully self-explanatory, natural deduction system. We begin with:

- | | | | |
|---|-----|---|-----|
| 1 | (1) | $\mathcal{D}^{999,999} \mathcal{D} \neg \mathcal{D}B1,000,000$ | (O) |
| 2 | (2) | $\mathcal{D}^{999,999} \neg \exists x (\mathcal{D} \mathcal{D}Bx \wedge \mathcal{D} \neg \mathcal{D}Bx')$ | (R) |

3	(3)	$\mathcal{D}\neg\mathcal{D}B1,000,000$	A
4	(4)	$\neg\exists x(\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg\mathcal{D}Bx')$	A
5	(5)	$\mathcal{D}\mathcal{D}B999,999$	A
3,5	(6)	$\mathcal{D}\mathcal{D}B999,999 \wedge \mathcal{D}\neg\mathcal{D}B1,000,000$	5,3 \wedge -I
3,5	(7)	$\exists x(\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg\mathcal{D}Bx')$	6 \exists -I
3,4,5	(8)	\perp	7,4 \neg -E
3,4	(9)	$\neg\mathcal{D}\mathcal{D}B999,999$	5,8 \neg -I
1,2	(10)	$\mathcal{D}^{999,999}\neg\mathcal{D}\mathcal{D}B999,999$	1,2,3,4,9 \mathcal{D} -C,

where the step from (9) to (10) employs the closure rule \mathcal{D} -C corresponding to (C) (note that (C) as well as (F) are crucial for (O) to entail (1)). With a qualification I will presently address, I assume that none of the inference rules used in the argument can plausibly be rejected. (R) justifies (2) if it is assumed, with the dominant approach, that (second-order) vagueness (definitely^{999,999}) requires (second-order) borderlineness.

Next, we have:

1,2	(1')	$\mathcal{D}^{999,998}\mathcal{D}\neg\mathcal{D}\mathcal{D}B999,999$	(10)
2'	(2')	$\mathcal{D}^{999,998}\neg\exists x(\mathcal{D}\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg\mathcal{D}\mathcal{D}Bx')$	(R)
3'	(3')	$\mathcal{D}\neg\mathcal{D}\mathcal{D}B999,999$	A
4'	(4')	$\neg\exists x(\mathcal{D}\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg\mathcal{D}\mathcal{D}Bx')$	A
5'	(5')	$\mathcal{D}\mathcal{D}\mathcal{D}B999,998$	A
3',5'	(6')	$\mathcal{D}\mathcal{D}\mathcal{D}B999,998 \wedge \mathcal{D}\neg\mathcal{D}\mathcal{D}B999,999$	5',3' \wedge -I
3',5'	(7')	$\exists x(\mathcal{D}\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg\mathcal{D}\mathcal{D}Bx')$	6' \exists -I
3',4',5'	(8')	\perp	7',4' \neg -E
3',4'	(9')	$\neg\mathcal{D}\mathcal{D}\mathcal{D}B999,998$	5',8' \neg -I
1,2,2'	(10')	$\mathcal{D}^{999,998}\neg\mathcal{D}\mathcal{D}\mathcal{D}B999,998$	1',2',3',4',9' \mathcal{D} -C.

(R) justifies (2') if it is assumed, with the dominant approach, that (third-order) vagueness (definitely^{999,998}) requires (third-order) borderlineness. But

999,998 structurally identical arguments eventually lead to ‘ $\neg\mathcal{D}^{1,000,001}B0$ ’, thereby contradicting (O).⁸

3.4.2 Weakening the Logic?

The only problematic inference used in the argument is *reductio* (in the steps from (8) to (9) and from (8′) to (9′)), which, whilst both classical and intuitionistically valid, is invalid in most many-valued logics which have seriously been proposed in the vagueness debate.⁹ *Reductio* has already crucially been used in the last step of the proof of theorem 3.2.1. There, we could have appealed instead to *weak contraposition*:

(WC) If $\Gamma, \varphi \vdash \psi$, then $\Gamma, \neg\psi \vdash \neg\mathcal{D}\varphi$

(which is valid e.g. in \mathbf{K}_3 on the most obvious definition of the truth conditions for ‘ \mathcal{D} ’ and should be valid, I submit, on any reasonable semantics for the operator). Letting ψ be ‘ $\exists x(\neg\mathcal{D}Bx \wedge \neg\mathcal{D}\neg Bx)$ ’, Γ be $\{(T), (M)\}$ and φ

⁸In conversation, Patrick Greenough has indicated that a structurally analogous paradox can be produced for *inexact knowledge* rather than *vagueness* (see section 6.4.5 for an explanation of the distinction). For concreteness, let the numerals denote the relevant pairwise indiscriminable trees in a series going from 25 ft tall to 15 ft tall, let ‘ \mathcal{D} ’ be interpreted as ‘An ordinary human subject can know by unaided observation that’ and let ‘ $B\xi$ ’ be interpreted as ‘ ξ is at least 20 ft tall’. Under this reading, (O) and the iterated “definitizations” of the relevant (first- and higher-order) borderlineness principles are still extremely plausible, so that a paradox structurally analogous to the higher-order sorites paradox just presented would arise. If I were to believe in borderlineness, I would block the higher-order sorites paradox using one of the logics to be developed in chapter 4, suitably extended with a ‘definitely’-operator. Crucially, such a solution could not be applied to this paradox of inexact knowledge, since, for all intents and purposes, all the expressions occurring in it may be taken to be precise, and so no vagueness-motivated weakening of the logic could be envisaged. It is my view that the solution to this latter paradox depends on understanding certain crucial features of the nature of reflective knowledge and of closure principles for knowledge. I try to make a start on these issues in Zardini [2007d].

⁹Thanks to Hartry Field for pressing this worry.

be ' $\exists x(\mathcal{D}Bx \wedge \mathcal{D}\neg Bx')$ ', using (WC) we can obtain the *weak-indefiniteness* principle:

$$(WI) \quad \neg \mathcal{D}\exists x(\mathcal{D}Bx \wedge \mathcal{D}\neg Bx').$$

Theorem 3.4.1. *Under minimal assumptions concerning \vdash , (WC) is equivalent with weak reductio:*

$$(WR) \quad \text{If } \Gamma, \varphi \vdash \perp, \text{ then } \Gamma \vdash \neg \mathcal{D}\varphi.^{10}$$

Proof.

- (WR) *entails* (WC). Suppose that $\Gamma, \varphi \vdash \psi$. Then, assuming the usual interaction between \neg and \perp , $\Gamma, \varphi, \neg\psi \vdash \perp$, and so, by (WR), $\Gamma, \neg\psi \vdash \neg \mathcal{D}\varphi$.
- (WC) *entails* (WR). Suppose that $\Gamma, \varphi \vdash \perp$. Then, by (WC), $\Gamma, \neg\perp \vdash \neg \mathcal{D}\varphi$, and so, by suppression of logical truths, $\Gamma \vdash \neg \mathcal{D}\varphi$.

□

With (WI) in place, we can run the following variation of the argument. We begin with:

1	(1)	$\mathcal{D}^{1,999,998}\mathcal{D}\mathcal{D}\neg \mathcal{D}B1,000,000$	(O)
2	(2)	$\mathcal{D}^{1,999,998}\neg \mathcal{D}\exists x(\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg \mathcal{D}Bx')$	(R)
3	(3)	$\mathcal{D}\mathcal{D}\neg \mathcal{D}B1,000,000$	A
4	(4)	$\neg \mathcal{D}\exists x(\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg \mathcal{D}Bx')$	A
5	(5)	$\mathcal{D}\mathcal{D}\mathcal{D}B999,999$	A
6	(6)	$\mathcal{D}\neg \mathcal{D}B1,000,000$	A
7	(7)	$\mathcal{D}\mathcal{D}B999,999$	A
6,7	(8)	$\mathcal{D}\mathcal{D}B999,999 \wedge \mathcal{D}\neg \mathcal{D}B1,000,000$	7,6 \wedge -I
6,7	(9)	$\exists x(\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg \mathcal{D}Bx')$	8 \exists -I
3,5	(10)	$\mathcal{D}\exists x(\mathcal{D}\mathcal{D}Bx \wedge \mathcal{D}\neg \mathcal{D}Bx')$	3,5,6,7,9 \mathcal{D} -C

¹⁰Thanks to Patrick Greenough for impressing upon me the importance of weak *reductio*.

3,4,5	(11)	\perp	10,4	\neg -E
3,4	(12)	$\neg \mathcal{D}\mathcal{D}\mathcal{D}\mathcal{D}B999, 999$	5,11	W \neg -I
1,2	(13)	$\mathcal{D}^{1,999,998}\neg \mathcal{D}\mathcal{D}\mathcal{D}\mathcal{D}B999, 999$	1,2,3,4,12	\mathcal{D} -C,

where the step from (11) to (12) employs the weak-*reductio* rule W \neg -I corresponding to (WR). (R) justifies (2) if it is assumed, with the dominant approach, that (second-order) vagueness (definitely^{1,999,998}) requires (second-order) borderlineness.

Next, we have:

1'	(1')	$\mathcal{D}^{1,999,996}\mathcal{D}\mathcal{D}\neg \mathcal{D}^4B999, 999$	(13)
2'	(2')	$\mathcal{D}^{1,999,996}\neg \mathcal{D}\exists x(\mathcal{D}\mathcal{D}^4Bx \wedge \mathcal{D}\neg \mathcal{D}^4Bx')$	(R)
3'	(3')	$\mathcal{D}\mathcal{D}\neg \mathcal{D}^4B999, 999$	A
4'	(4')	$\neg \mathcal{D}\exists x(\mathcal{D}\mathcal{D}^4Bx \wedge \mathcal{D}\neg \mathcal{D}^4Bx')$	A
5'	(5')	$\mathcal{D}\mathcal{D}\mathcal{D}^4B999, 998$	A
6'	(6')	$\mathcal{D}\neg \mathcal{D}^4B999, 999$	A
7'	(7')	$\mathcal{D}\mathcal{D}^4B999, 998$	A
6',7'	(8')	$\mathcal{D}\mathcal{D}^4B999, 998 \wedge \mathcal{D}\neg \mathcal{D}^4B999, 999$	7',6' \wedge -I
6',7'	(9')	$\exists x(\mathcal{D}\mathcal{D}^4Bx \wedge \mathcal{D}\neg \mathcal{D}^4Bx')$	8' \exists -I
3',5'	(10')	$\mathcal{D}\exists x(\mathcal{D}\mathcal{D}^4Bx \wedge \mathcal{D}\neg \mathcal{D}^4Bx')$	3',5',6',7',9' \mathcal{D} -C
3',4',5'	(11')	\perp	10',4' \neg -E
3',4'	(12')	$\neg \mathcal{D}\mathcal{D}\mathcal{D}\mathcal{D}^4B999, 998$	5',11' W \neg -I
1',2'	(13')	$\mathcal{D}^{1,999,996}\neg \mathcal{D}\mathcal{D}\mathcal{D}\mathcal{D}^4B999, 998$	1',2',3',4',12' \mathcal{D} -C,

(R) justifies (2') if it is assumed, with the dominant approach, that (fifth-order) vagueness (definitely^{1,999,996}) requires (fifth-order) borderlineness. But 999,998 structurally identical arguments eventually lead to ' $\neg \mathcal{D}^{3,000,001}B0$ ', thereby contradicting (O).

Might not a defender of the dominant approach weaken the modality and only maintain the weaker claim that, if an expression is vague, it is not definitely the case that it does not present borderline cases (instead

of the usual, stronger claim that, if an expression is vague, it presents borderline cases)? No. For $\exists x(\mathcal{D}Bx \wedge \mathcal{D}\neg Bx')$ uncontroversially entails $\neg\exists x(\neg\mathcal{D}Bx \wedge \neg\mathcal{D}\neg Bx)$, and so, by (C), $\mathcal{D}\exists x(\mathcal{D}Bx \wedge \mathcal{D}\neg Bx')$ entails $\mathcal{D}\neg\exists x(\neg\mathcal{D}Bx \wedge \neg\mathcal{D}\neg Bx)$, wherefore, by (WC), $\neg\mathcal{D}\neg\exists x(\neg\mathcal{D}Bx \wedge \neg\mathcal{D}\neg Bx)$ entails $\neg\mathcal{D}\mathcal{D}\exists x(\mathcal{D}Bx \wedge \mathcal{D}\neg Bx')$, which, by a slight modification of the argument just presented, can also be shown to be soritical.¹¹

In closing, let me stress that inspection of our higher-order sorites paradox immediately reveals that the transfinite strength of (O) and (R) is wholly dispensable. For once the finite length of the relevant soritical series¹² has been fixed, only weakenings of (O) and (R) where \mathcal{D}^ω is replaced by sufficiently large finite concatenations of \mathcal{D} are needed to breed paradox. Even if the considerations adduced in support of (O) and (R) should prove resistible, the defender of the dominant approach is still faced with the daunting task of showing that no suitable finite weakenings of (O) and (R) are ever available for any soritical series—that indisputable positive and negative cases are much less definitely so than we thought them to be and that higher-order vagueness is much less definitely radical than we thought it to be. I for one cannot see how the defender of the dominant approach has in the present state of information a guarantee that that task can be accomplished in its full generality.^{13,14}

3.5 Conclusion

Our paradox seems then to show higher-order principles of the form of (I) (and, therefore, higher-order principles of the form of (B)), put forward by

¹¹Thanks to Hartry Field here.

¹²A *soritical series* is any series of objects which, like S , can induce a sorites paradox.

¹³Thanks to Crispin Wright for emphasizing to me the dialectical importance of this point.

¹⁴Further paradoxes of higher-order vagueness for the dominant approach are developed in Zardini [2006a]; Zardini [2006b]. Greenough [2005]; Wright [2007b] are the best up-to-date discussions of these issues.

the dominant approach in order to characterize, for every i , the nature of the vagueness of ' $\mathcal{D}^i B$ ', to be inconsistent with (O). The situation is thus fully analogous to the one faced by the naive theory of vagueness, where the original sorites paradox apparently showed (TOL), put forward by the naive theory in order to characterize the nature of the vagueness of ' B ', to be inconsistent with 0's being B and 1,000,000's being $\neg B$. Hence, the dominant approach loses its main advantage against the naive theory—its alleged immunity to any form of sorites paradox. A new diagnosis of the paradox is called for. To this we turn in the next chapter.

Chapter 4

Tolerant Logics

4.1 Introduction and Overview

Some men are bald: a man with absolutely no hairs on his scalp is bald. Other men are not bald: a man with a full head of hairs is not bald. It is also very intuitive that the difference between the bald and the non-bald is not an exquisitely fine difference: if a man is bald, so is one with just one more hair on his scalp. In other words, it is also very intuitive that the expression ‘bald’ is tolerant, in the sense that one-hair differences do not make a difference to its positive or negative application (as in chapter 3, I will acquiesce in the usual, harmlessly false assumption that baldness is just a matter of the number of hairs on one’s scalp). As we have seen in the previous chapters, the theory based on the union of these or relevantly similar claims may well be called ‘the naive theory of vagueness’. It is a *theory of vagueness* insofar as its trio of claims can be taken to give an account of what the vagueness of an expression consists in—very roughly, in its being tolerant whilst having both positive and negative cases of application (see section 1.4 for a more careful formulation). It is a *naive* theory of vagueness insofar as its *prima facie* theoretical advantages of simplicity (chapter 1), explanatory power (chapter 1) and preservation of ordinary intuitions (chapter 2) are

cast into grave doubt by an argument—the so-called “*sorites paradox*”—purporting to show its inconsistency, in a way similar to that in which the naive theory of sets, based on the unrestricted set-comprehension schema, is affected by the set-theoretical paradoxes.

An informal presentation of the sorites paradox goes as follows. Consider the premises:

- (1) A man with 0 hairs is bald;
- (2) A man with 1,000,000 hairs is not bald;
- (3) If a man with i hairs is bald, so is a man with $i + 1$ hairs.

The naive theory of vagueness is committed to all these three premises. However, from (3) we have that, if a man with 0 hairs is bald, so is a man with 1 hair, which, together with (1), yields that a man with 1 hair is bald. Yet, from (3) we also have that, if a man with 1 hair is bald, so is a man with 2 hairs, which, together with the previous lemma that a man with 1 hair is bald, yields that a man with 2 hairs is bald. With another 999,997 structurally identical arguments, we reach the conclusion that a man with 999,999 hairs is bald. From (3) we also have that, if a man with 999,999 hairs is bald, so is a man with 1,000,000 hairs, which, together with the previous lemma that a man with 999,999 hairs is bald, yields that a man with 1,000,000 hairs is bald. It would then seem that the contradictory of (2) follows simply from (1) and (3).¹

I think that all the three characteristic claims of the naive theory of vagueness are *in fact* true and jointly provide a satisfactory account of what the vagueness of an expression consists in, but I will not argue directly for

¹In the simple form presented in the text, the argument has to make use of 999,999 applications of the *contraction* rule as well (see Copeland [1997], pp. 523–6). However, clumsier versions of it are readily available which only employ 1,000,000 different instances of (3) instead of (3) itself, and so do not require contraction (against what Copeland seems to imply).

that in this chapter. Rather, I will focus here on the problem presented by the sorites paradox to the naive theory of vagueness, offering the basics of a logical framework in which the theory *can* be true. As we will see, the basic idea of the logical framework is not that of weakening the properties of any particular logical constant (i.e. weakening the *operational rules* of the logic). While this might be desirable in other respects, it won't help much in saving the naive theory of vagueness from inconsistency, since, as we have seen, the sorites paradox need only use the rule of *modus ponens* for the conditional used to formulate the tolerance principles and rejection of this rule—in addition of course to being in itself highly implausible—would seem to deprive those principles of their intended *force* (which consists in *not allowing any difference* in, say, baldness between two men who only differ of 1 in the number of hairs on their scalp). The basic idea of the logical framework is rather that of weakening the properties of the consequence relation itself (i.e. weakening the *structural rules* of the logic), in particular the *transitivity* property (see chapter 5 for an extended philosophical discussion of both the viability and import of non-transitivist proposals). I will call any logic in which the naive theory of vagueness is consistent owing to the failure of the transitivity property 'a *tolerant logic*'.

No doubt vagueness has offered other reasons, apparently independent of the sorites paradox, to question the validity of certain classical laws and rules concerning specific logical constants: the phenomenon of borderlineness—especially once it is interpreted as strong borderlineness—has for instance induced many a commentator to reject the law of excluded middle (see e.g. Field [2003a]). I will thus start by describing a very minimal logical basis which, while being non-transitive, does not prejudge any of the other many interesting issues arising in the philosophy of the logic of vagueness. From such a basis, I will then build up to my favoured specific system (mirroring, in a non-transitive setting, classical logic), indicating along the way how other weaker non-transitive systems (mirroring, in a non-transitive setting, some or other non-classical logic) can be defined. An appropriate justification of

such preference lies outside the scope of this essay: suffice it to say that I regard classical patterns of reasoning as unchallenged by vagueness insofar as they are compatible with a suitable weakening of the transitivity of the consequence relation (for e.g. a defense of the law of excluded middle see Zardini [2007a]).

The framework is thus *neutral* in the sense that it can fruitfully be employed also by a theorist who, while attracted to the naive theory of vagueness, has also reason to reject some other classical pattern of reasoning for vague expressions. Moreover, as shown in chapter 3, most contemporary theories of vagueness, requiring strong borderline cases for an expression whenever the expression is vague, are also subject to (higher-order) sorites paradoxes. These paradoxes too can be blocked in a principled way by restricting the transitivity of the consequence relation. Therefore, such theories too should find the present framework congenial: its neutrality will allow them to focus on (one of) the particular logic(s) which is generated by restricting the transitivity of their originally favoured consequence relation.

The rest of the chapter is organized as follows. Section 2 provides and explains the general many-valued semantic framework of tolerant logics. Building on this basic framework, section 3 stepwise develops a tolerant counterpart to classical zeroth-order logic and investigates some of its properties. Section 4 discusses how the notion of consistency is best modelled in the present framework and offers a proof of the consistency in a tolerant logic of the zeroth-order fragment of the naive theory of vagueness. Section 5 extends the semantic construction to a first-order language with identity, offering a corresponding consistency proof and individuating a specific tolerant logic as being on balance the best logic for a vague language. Section 6 draws the conclusions which follow for the status of the naive theory from the non-transitivist solution to the sorites paradox and hints at some of the outstanding technical issues.

4.2 A Neutral Framework

4.2.1 Syntax

The core idea of tolerant logics—the failure of the transitivity property for the consequence relation—already emerges at the sentential level. Furthermore, a significant fragment of the naive theory of vagueness can be adequately regimented in a standard sentential language. Hence, in this and the two following sections, we will focus on the study of such a language, the vague zeroth-order language \mathcal{L}^0 . To stress, it is not that \mathcal{L}^0 is said to be vague because some of its expressions are vague: being a formal language, its non-logical atomic expressions are not interpreted, and so *a fortiori* not vague. Moreover, there is no plausibility in the idea that the mathematically precise semantics which will be provided for its logical expressions should induce any vagueness in them. Rather, \mathcal{L}^0 is said to be vague because the logic determined by its semantics is such that an adequate regimentation of a significant fragment of the naive theory of vagueness is consistent in it. I should also emphasize that the *metalanguage* \mathcal{M} within which we will conduct our study of \mathcal{L}^0 will be assumed to be classical (in particular, assumed to be such that its consequence relation is transitive), and that the *metatheory* used in \mathcal{M} will be the classical set theory ZFC.²

Definition 4.2.1. The set $AS_{\mathcal{L}^0}$ of the *atomic symbols* of \mathcal{L}^0 is defined by enumeration as follows:³

- The denumerable set $VAR_{\mathcal{L}^0}$ of variables $P_0, P_1, P_2 \dots, Q_0, Q_1, Q_2 \dots, R_0, R_1, R_2 \dots$ is a subset of $AS_{\mathcal{L}^0}$;

²The use of a classical metalanguage in the explanation of a non-classical object language is of course one of the cruxes of any proposal of deviation from classical logic. I cannot hope to address here the philosophical issues related to this asymmetry nor the crucial philosophical and technical question as to whether and how much of an explanation of a non-classical object language can be provided using a metalanguage with the same logic.

³Throughout, formal symbols are understood autonomously to refer to themselves.

- The 1ary sentential operator \neg belongs to $AS_{\mathcal{L}^0}$;
- The 2ary sentential operators \wedge , \vee and \rightarrow belong to $AS_{\mathcal{L}^0}$;
- (and) belong to $AS_{\mathcal{L}^0}$;
- Nothing else belongs to $AS_{\mathcal{L}^0}$.

Note that, given the generality of the framework we will be developing, we cannot assume the definability of any of the standard sentential operators in terms of others. In particular, in order to let non-transitivity emerge already at the very first stages of the construction, it will prove useful to have a primitive conditional \rightarrow bearing a privileged connection to the consequence relation.

Definition 4.2.2. The set $WFF_{\mathcal{L}^0}$ of *well-formed formulae* (wffs) of \mathcal{L}^0 (equivalently, given that \mathcal{L}^0 is zeroth-order, the set of sentences of \mathcal{L}^0) can be defined by recursion in the usual way:⁴

- If $\varphi \in VAR_{\mathcal{L}^0}$, $\varphi \in WFF_{\mathcal{L}^0}$;
- If $\varphi \in WFF_{\mathcal{L}^0}$, $(\neg\varphi) \in WFF_{\mathcal{L}^0}$;
- If $\varphi, \psi \in WFF_{\mathcal{L}^0}$, $(\varphi \wedge \psi), (\varphi \vee \psi), (\varphi \rightarrow \psi) \in WFF_{\mathcal{L}^0}$;
- Nothing else belongs to $WFF_{\mathcal{L}^0}$.

Henceforth, to save on brackets, I will assume the usual scope hierarchy among the operators (with \neg binding more strongly than \wedge and \vee , and with these in turn binding more strongly than \rightarrow) and left associativity of the 2ary operators (so that $(\varphi_0 \star \varphi_1 \star \varphi_2 \dots \star \varphi_i)$ reads $((\dots(\varphi_0 \star \varphi_1) \star \varphi_2) \dots \star \varphi_i)$, with \star being a 2ary operator). I will also drop the outermost brackets of a self-standing wff.

⁴Throughout, ‘ φ ’, ‘ ψ ’ and ‘ χ ’ (possibly with numerical subscripts) are used as metalinguistic variables ranging over $WFF_{\mathcal{L}^0}$.

4.2.2 Tolerant Semantic Structures

Definition 4.2.3. We will take a *sequence* of wffs of \mathcal{L}^0 to be a function whose domain is some suitable initial segment of the ordinals and whose range is a subset of $WFF_{\mathcal{L}^0}$.⁵ As usual:⁶

- $\Gamma, \Delta := \Gamma \cup \{\langle \delta_1, \varphi \rangle : \text{for some } \delta_0 \in \text{dom}(\Delta), \delta_1 = \text{dom}(\Gamma) + \delta_0 \text{ and } \varphi = \Delta(\delta_0)\}$;
- $\Gamma, \varphi := \Gamma, \langle \varphi \rangle$,

where ‘:=’ is a metalinguistic symbol expressing the definition relation and $\text{dom}(R)$ and $\text{ran}(R)$ are, respectively, the domain and the range of R .

Definition 4.2.4. A *logic* \mathbf{L} for \mathcal{L}^0 is any subset of $\{\langle \Gamma_0, \Gamma_1 \rangle : \text{ran}(\Gamma_0), \text{ran}(\Gamma_1) \subseteq WFF_{\mathcal{L}^0}\}$.⁷

As anticipated, we start our semantic construction by developing a minimal logical basis, the basic tolerant logic \mathbf{T}^0 , which already displays the core idea of any tolerant logic—the failure of the transitivity property for the consequence relation—but which is otherwise completely neutral with regard to the issues concerning the philosophy of the logic of a vague language. By adding further and further constraints on the semantics, we will then be able to specify stronger and stronger tolerant logics. We will make use of standard lattice-theoretical semantics, introducing the modifications appropriate for obtaining failures of the transitivity property for the consequence relation.

Definition 4.2.5. A \mathbf{T}^0 -*structure* \mathfrak{S} for \mathcal{L}^0 is a 5ple $\langle V_{\mathfrak{S}}, D_{\mathfrak{S}}, \preceq_{\mathfrak{S}}, \text{tol}_{\mathfrak{S}}, O_{\mathfrak{S}} \rangle$, where:

⁵Every function is assumed to be total unless otherwise specified.

⁶Throughout, ‘ Γ ’, ‘ Δ ’, ‘ Θ ’, ‘ Λ ’, ‘ Ξ ’ (possibly with numerical subscripts) are used as metalinguistic variables ranging over the set of sequences whose range is included in $WFF_{\mathcal{L}^0}$.

⁷We will thus be working in a multiple-conclusion framework, this being required in order to achieve the desired generality.

- $V_{\mathfrak{S}}$ is a non-empty set of objects (the “values”);
- $D_{\mathfrak{S}}$ is a non-empty subset of $V_{\mathfrak{S}}$ (the “designated values”) such that:

(D_0) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $v_0 \in D_{\mathfrak{S}}$ and $v_0 \preceq_{\mathfrak{S}} v_1$, then $v_1 \in D_{\mathfrak{S}}$
(see the next item for a definition of $\preceq_{\mathfrak{S}}$)

($D_{\mathfrak{S}}$ is an *upper set*);

- $\preceq_{\mathfrak{S}}$ is a partial ordering (reflexive, anti-symmetric, transitive relation) on $V_{\mathfrak{S}}$ such that:

(glb/lub₀²) For every $v_0, v_1 \in V_{\mathfrak{S}}$, $\{v_0, v_1\}$ has a greatest lower bound (glb) and a least upper bound (lub)

($\preceq_{\mathfrak{S}}$ thus corresponds to a *lattice*);

- $\text{tol}_{\mathfrak{S}}$ is a “tolerance” function from $V_{\mathfrak{S}}$ into $\text{pow}(V_{\mathfrak{S}})$ (the powerset of $V_{\mathfrak{S}}$) such that:

(tol_0) For every $v \in V_{\mathfrak{S}}$, $v \in \text{tol}_{\mathfrak{S}}(v)$;

(tol_1) For every $v \in V_{\mathfrak{S}}$, $\text{tol}_{\mathfrak{S}}(v)$ is an upper set.

Note in particular that $\text{tol}_{\mathfrak{S}}(v)$ is allowed to contain values which are *not* contained in the upper set whose minimum element is v . As we will see, this “tolerating” feature of tol is crucial in generating failures of the transitivity property for the consequence relation;

- $O_{\mathfrak{S}}$ is a non-empty set of operations on $V_{\mathfrak{S}}$. In particular, $\{\text{neg}_{\mathfrak{S}}, \text{impl}_{\mathfrak{S}}\} \subseteq O_{\mathfrak{S}}$, where:

($\text{neg}_0^{\rightarrow}$) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $v_0 \preceq_{\mathfrak{S}} v_1$, then $\text{neg}_{\mathfrak{S}}(v_1) \preceq_{\mathfrak{S}} \text{neg}_{\mathfrak{S}}(v_0)$;⁸

⁸Note that ($\text{neg}_0^{\rightarrow}$) and ($\text{neg}_1^{\rightarrow}$) can be neatly packaged into the so-called “*law of intuitionist contraposition*” that, for every $v_0, v_1 \in V_{\mathfrak{S}}$, if $v_0 \preceq_{\mathfrak{S}} \text{neg}_{\mathfrak{S}}(v_1)$, then $v_1 \preceq_{\mathfrak{S}} \text{neg}_{\mathfrak{S}}(v_0)$.

- ($\text{neg}_1^{\rightarrow}$) For every $v \in V_{\mathfrak{S}}$, $v \preceq_{\mathfrak{S}} \text{neg}_{\mathfrak{S}}(\text{neg}_{\mathfrak{S}}(v))$;
- ($\text{impl}_0^{\rightarrow}$) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $v_1 \in \text{tol}_{\mathfrak{S}}(v_0)$, then $\text{impl}_{\mathfrak{S}}(v_0, v_1) \in D_{\mathfrak{S}}$;
- ($\text{impl}_0^{\leftarrow}$) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $\text{impl}_{\mathfrak{S}}(v_0, v_1) \in D_{\mathfrak{S}}$, then $v_1 \in \text{tol}_{\mathfrak{S}}(v_0)$.

Again, note in particular how impl relates to tol , and especially how ($\text{impl}_0^{\leftarrow}$) allows $\text{impl}_{\mathfrak{S}}(v_0, v_1)$ to belong to $D_{\mathfrak{S}}$ even if v_1 does not belong to the upper set whose minimum element is v_0 . Henceforth, we will focus on the case where $\{\text{neg}_{\mathfrak{S}}, \text{impl}_{\mathfrak{S}}\} = O_{\mathfrak{S}}$, but it is clear how, given the rich structure generated by $V_{\mathfrak{S}}$, $D_{\mathfrak{S}}$, $\preceq_{\mathfrak{S}}$ and $\text{tol}_{\mathfrak{S}}$, many other interesting operations may be defined and added to $O_{\mathfrak{S}}$ (and be expressed by corresponding operators in some extension of \mathcal{L}^0).⁹ In order to exploit the full power of ($\text{neg}_0^{\rightarrow}$), we place another constraint on $\text{tol}_{\mathfrak{S}}$:

- (tol_2) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $v_1 \in \text{tol}_{\mathfrak{S}}(v_0)$, then $\text{neg}_{\mathfrak{S}}(v_0) \in \text{tol}_{\mathfrak{S}}(\text{neg}_{\mathfrak{S}}(v_1))$.

A \mathbf{T}^0 -structure can then be used to interpret \mathcal{L}^0 once it is equipped with an interpretation function for $WFF_{\mathcal{L}^0}$ and once suitable recursive clauses for the sentential operators are given.

⁹In particular, given the characteristic “lowering” behaviour of the conjunction operation (to be specified shortly), it may prove useful to define an equivalence operation $\text{equiv}_{\mathfrak{S}}$ such that:

- ($\text{equiv}_0^{\rightarrow}$) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $v_1 \in \text{tol}_{\mathfrak{S}}(v_0)$ and $v_0 \in \text{tol}_{\mathfrak{S}}(v_1)$, then $\text{equiv}_{\mathfrak{S}}(v_0, v_1) \in D_{\mathfrak{S}}$;
- ($\text{equiv}_0^{\leftarrow}$) For every $v_0, v_1 \in V_{\mathfrak{S}}$, if $\text{equiv}_{\mathfrak{S}}(v_0, v_1) \in D_{\mathfrak{S}}$, then $v_1 \in \text{tol}_{\mathfrak{S}}(v_0)$ and $v_0 \in \text{tol}_{\mathfrak{S}}(v_1)$,

and extend \mathcal{L}^0 with a new primitive, biconditional-like, 2ary operator \leftrightarrow expressing $\text{equiv}_{\mathfrak{S}}$. Save for a brief remark later (fn 18) substantiating this point, in this essay I will not pursue further the investigation of the logic of such \leftrightarrow .

Definition 4.2.6. A \mathbf{T}^0 -model \mathfrak{M} for \mathcal{L}^0 based on a \mathbf{T}^0 -structure \mathfrak{S} is a 6ple $\langle V_{\mathfrak{M}}, D_{\mathfrak{M}}, \preceq_{\mathfrak{M}}, \text{tol}_{\mathfrak{M}}, O_{\mathfrak{M}}, \text{int}_{\mathfrak{M}} \rangle$, where $V_{\mathfrak{M}}$, $D_{\mathfrak{M}}$, $\preceq_{\mathfrak{M}}$, $\text{tol}_{\mathfrak{M}}$ and $O_{\mathfrak{M}}$ are identical to $V_{\mathfrak{S}}$, $D_{\mathfrak{S}}$, $\preceq_{\mathfrak{S}}$, $\text{tol}_{\mathfrak{S}}$ and $O_{\mathfrak{S}}$ respectively, and $\text{int}_{\mathfrak{M}} : VAR_{\mathcal{L}^0} \mapsto V_{\mathfrak{M}}$ is an interpretation function for $VAR_{\mathcal{L}^0}$.

Definition 4.2.7. $\text{int}_{\mathfrak{M}}$ is extended to a full valuation function $\text{val}_{\mathfrak{M}} : WFF_{\mathcal{L}^0} \mapsto V_{\mathfrak{M}}$ by the following recursion:

- $(\text{val}_{VAR_{\mathcal{L}^0}})$ If $\varphi \in VAR_{\mathcal{L}^0}$, $\text{val}_{\mathfrak{M}}(\varphi) = \text{int}_{\mathfrak{M}}(\varphi)$;
- (val_{\neg}) $\text{val}_{\mathfrak{M}}(\neg\varphi) = \text{neg}_{\mathfrak{M}}(\text{val}_{\mathfrak{M}}(\varphi))$;
- (val_{\wedge}) $\text{val}_{\mathfrak{M}}(\varphi \wedge \psi) = \text{glb}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}}(\varphi), \text{val}_{\mathfrak{M}}(\psi)\})$;
- (val_{\vee}) $\text{val}_{\mathfrak{M}}(\varphi \vee \psi) = \text{lub}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}}(\varphi), \text{val}_{\mathfrak{M}}(\psi)\})$;
- $(\text{val}_{\rightarrow})$ $\text{val}_{\mathfrak{M}}(\varphi \rightarrow \psi) = \text{impl}_{\mathfrak{M}}(\text{val}_{\mathfrak{M}}(\varphi), \text{val}_{\mathfrak{M}}(\psi))$.

Determining which value $v \in V_{\mathfrak{M}}$ a wff has in \mathfrak{M} , $\text{val}_{\mathfrak{M}}$ *a fortiori* determines whether $v \in D_{\mathfrak{M}}$ or not—in other words, $\text{val}_{\mathfrak{M}}$ determines whether the value of a wff is designated or not. Now, in standard many-valued semantics, the role played by designated values can be (informally) explained as follows. It is assumed that the *actual* semantics of an interpreted language \mathcal{J} whose logical properties one is interested in studying exhibits at least the *general features* of the semantics used in the mathematical study of a formal language \mathcal{K} into which \mathcal{J} can be adequately regimented. For example, \mathcal{J} might be a fragment of English expressing first-order Peano-Dedekind arithmetic and \mathcal{K} be a standard formal first-order language (with identity and functors): then just as, in every model of \mathcal{K} , every sentence (closed wff) of \mathcal{K} is assigned either 1 or 0 as value (but not both),¹⁰ so it is assumed that every sentence of \mathcal{J} is either true or false (but not both). If the formal semantics of \mathcal{K}

¹⁰Of course, even if usual, the particular choice of 1 and 0 to model truth and falsity is completely conventional—any two other objects recognized by the background mathematical theory will do.

is many-valued, there will typically be, included in the set of all the values a sentence can be assigned in a model, a set of designated values. What do such values correspond to in the semantics of \mathcal{J} ?

The usual answer in philosophy of logic to this question is, roughly, that they correspond to the “*good*” values a sentence can have—that is, those which make a sentence good enough to be asserted, good enough to be believed, good enough to be acted upon etc. (see Priest [2001], p. 216).¹¹ At least in this respect, then, what designated values correspond to plays the same role in a many-valued framework as *truth* plays in a two-valued framework, for, in a two-valued framework, it is truth that which warrants assertion, belief and action.

This is the place to enter a crucial clarification concerning the present use of a many-valued semantics.¹² Such a semantics is here used with the main purpose of inducing a certain (family of) logic(s). The different values are supposed to model the different levels of goodness a sentence can have in terms of its assertability, believability, enactability etc., where the notion of a level of goodness can be reduced for simple predications to the position of the relevant object in the ordering generated by the contextually relevant dimension of comparison. The values assigned to compound sentences by the semantics are supposed to model the way in which we understand the level of goodness of a compound sentence to be determined by the level of goodness of its components. Emphatically, the different intermediate levels of goodness are not different ways in which a sentence can be neither true nor false (and so neither do the extreme levels of goodness—if they exist—coincide with truth and falsity), nor do levels of goodness represent an ordering of truth among sentences (on this very last point I thus diverge from the interpreta-

¹¹ “*Semantically*” good enough. For it may well be that other non-semantic features (e.g. epistemic ones) contribute to the determination of a sentence’s assertability, believability, enactability etc. A similar qualification concerning truth should be understood below as implicit.

¹² Thanks to Graham Priest for urging consideration of this issue.

tion of the lattice-theoretical many-valued framework offered by Weatherson [2005a]). The truth about truth is that a sentence ‘ P ’ is true iff P , and false otherwise—and this is manifestly too simple a notion to use if one is trying to develop a non-classical logic in a classical metatheory. It is levels of goodness, understood in the minimalist way just explained, which allow us to draw the fine-grained distinctions required by non-classical reasoning (I thus agree with Michael Dummett’s strictures against those who “reduce the semantic notion of logical consequence to a purely algebraic tool” (Dummett [1975b], p. 293), but disagree with him when he claims that “[o]n the assumption that all our sentences possess determinate truth-values, there is simply nothing that one can think of that a truth-table would leave unexplained concerning the meaning of the sentential operator for which it was correct” (Dummett [1975b], p. 294)).

Let me note that understanding logical operations as operating on the fine structure of levels of goodness rather than simply on truth and falsity does not mean of course that they are not significantly constrained by truth and falsity—indeed, they should be such that the logic they generate is not too strong as to rule out intuitively possible assignments of truth and falsity and not too weak as to allow intuitively impossible assignments of truth and falsity. Let me also note that an additional layer of complexity is induced by the fact that talk of ‘good values’ should presumably be vague, so that our use of a classical metalanguage risks to misrepresent what good values are. This problem connects with some of the issues mentioned in fn 2—here I will only add that the risk is at least partially averted by the quantification over models in the definition of the consequence relation and by the rejection that one of the classically described models is the intended model of a vague language. Let me finally stress that I am painfully aware that a much more detailed discussion of these issues would be needed in order to make my claims persuasive, but that I hope that what I have just said gives enough indication as to how to understand the framework to be developed.

In a two-valued framework, truth also plays a crucial role in the definition

of the consequence relation: the informal characterization of consequence as *necessary truth preservation from the premises to the conclusions* gets formally translated as preservation of 1 from the premises to the conclusions in every model. In a many-valued framework, it is then very natural to define consequence as preservation of designated value from the premises to the conclusions (in the sense that, if every premise has a designated value in a model \mathfrak{M} , then some conclusion also has a designated value in \mathfrak{M}). It is thus guaranteed that, when one argues validly from good premises, one will reach some good conclusion.

This is the point of entry of the crucial modification I would like to propose in order to generate failures of the transitivity property for the consequence relation. Consider again the tolerance function tol :¹³ informally, it implements the idea that, if v_0 counts as *very good* a value for a sentence to have, any $v_1 \in \text{tol}(v_0)$ will also count as *good enough* a value for a sentence to have (to be asserted, believed, acted upon etc.). Of course, since we are working in a classical (and thus transitive) metalanguage, we cannot require that, if v_0 counts as *good enough* a value for a sentence to have, any $v_1 \in \text{tol}(v_0)$ will likewise count as *good enough* a value for a sentence to have, since such a principle would breed paradox (given the fact that it need not be the case that, for every $v_0, v_1, v_2 \in V$, if $v_2 \in \text{tol}(v_1)$ and $v_1 \in \text{tol}(v_0)$, then $v_2 \in \text{tol}(v_0)$).¹⁴ Still, it might reasonably be argued that our inferential practices with a vague language lend support to the idea that all consequence in a vague language guarantees is that, when one argues validly from very good premises, one will reach some good enough conclusion. I will not attempt here to establish this claim in its generality—rather, I will briefly try to support its plausibility by illustrating how it is supposed to work in a particular case.

¹³Henceforth, I will drop subscripts for models and structures if no ambiguity threatens.

¹⁴Letting $\text{tol}^* = \{\langle X, Y \rangle : X \subseteq V \text{ and, for every } v_0, v_0 \in Y \text{ iff, for some } v_1 \in X, v_0 \in \text{tol}(v_1)\}$, the point in the text can be put by saying that tol^* is *non-idempotent*, in the sense that it need not be the case that, for every $X \subseteq V$, $\text{tol}^*(\text{tol}^*(X)) = \text{tol}^*(X)$.

(4) A man with i hairs is bald

might have a very good value, because it has been accepted on the basis of perception, or intuition, or testimony etc.

(5) If a man with i hairs is bald, then a man with $i + 1$ is bald

might also have a very good value, because it belongs to the naive theory of vagueness. Given this, it is reasonable to expect that:

(6) A man with $i + 1$ hairs is bald,

which presumably follows from (5) and (4), will have a good enough value, and thus that it will also be assertable, believable, enactable etc. However, it is not equally reasonable to expect that, because:

(7) If a man with $i + 1$ hairs is bald, then a man with $i + 2$ is bald

has also a very good value (since it too belongs to the naive theory of vagueness),

(8) A man with $i + 2$ hairs is bald,

which presumably follows from (7) and (6), will also have a good enough value: for (6) only has a good enough value, and it is not equally reasonable to expect that a sentence having a merely good enough value can always be fed as a premise into an inferential process to yield a conclusion with a good enough value (as it is the case, on the contrary, when every premise has not just a good enough value, but a very good value).

It seems to me that the description of the case I have given agrees with those which would be given by many speakers if subjected to this short stretch of a soritical series. Typically, they would accept (4), (5) and (7) (and (6))

without feeling thereby compelled to accept (8), even though they would accept the validity of both *modus-ponens* arguments (note that, typically, they would not feel compelled to accept (8) even if there were no evidence that it is false). Typically, they would justify this complex pattern of acceptance by claiming that, even though in itself valid, the argument in question (i.e. *modus ponens*) should not be “pushed too far”. In order to explain these reactions, it is plausible to conjecture that the underlying implicit conception of validity is such that it is only thought to guarantee that very good premises yield a good enough conclusion, so that it need not be thought to guarantee that a double application of *modus ponens* starting from very good premises will ultimately issue in a good enough conclusion (as the case just described makes clear). On this conception—which is the key to the failure of transitivity—validity is not a matter of *preservation* of anything (neither of being very good from the premises to the conclusions nor of being good enough from the premises to the conclusions), but a matter of *connection* between the premises’ being very good and the conclusions’ being good enough (see section 5.3.2 for more on this point).

4.2.3 Basic Tolerant Logic

Now, our semantic structures give us all the resources to capture the distinction between very good values and good enough values and to deploy it in order to define a consequence relation whose hallmark is the guarantee that very good premises lead to good enough conclusions.

Definition 4.2.8. The set of “tolerated values” $T_{\mathfrak{S}}$ of a structure \mathfrak{S} is defined as follows:

$$(T) \quad T_{\mathfrak{S}} := \bigcup_{d \in D_{\mathfrak{S}}} \text{tol}(d)$$

We can then let the set of designated values $D_{\mathfrak{S}}$ represent the very good values of \mathfrak{S} , and let the set of tolerated values $T_{\mathfrak{S}}$ represent the good enough values of \mathfrak{S} .

Theorem 4.2.1. *T is an upper set.*

Proof. Suppose that $v_0 \in T$ and $v_0 \preceq v_1$. Then, for some $d \in D$, $v_0 \in \text{tol}(d)$, and so is such that, for every v_2 , if $v_0 \preceq v_2$, $v_2 \in \text{tol}(d)$ as well. Therefore, since $v_0 \preceq v_1$, $v_1 \in \text{tol}(d)$ as well, and so $v_1 \in T$ as well. □

Definition 4.2.9. The consequence relation on pairs of sequences of wffs $\in WFF_{\mathcal{L}^0}$ constituting the basic tolerant logic \mathbf{T}^0 ($\models_{\mathbf{T}^0}$) is defined as follows:

($\models_{\mathbf{T}^0}$) A sequence of wffs Δ is a $\models_{\mathbf{T}^0}$ -consequence of a sequence of wffs Γ ($\Gamma \models_{\mathbf{T}^0} \Delta$) iff, for every \mathbf{T}^0 -model \mathfrak{M} , if, for every $\varphi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\varphi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$.

It's easy to check that \mathbf{T}^0 exhibits some basic *structural properties* (note that full associativity is already guaranteed by our choice of sequences as terms of the consequence relation):

Theorem 4.2.2. *\mathbf{T}^0 exhibits the following structural properties.¹⁵*

(C^l) *If $\Gamma_0 \vdash \Delta$, then, for every Γ_1 such that $\text{fld}(\Gamma_0) = \text{fld}(\Gamma_1)$, $\Gamma_1 \vdash \Delta$;*

(C^r) *If $\Gamma \vdash \Delta_0$, then, for every Δ_1 such that $\text{fld}(\Delta_0) = \text{fld}(\Delta_1)$, $\Gamma \vdash \Delta_1$;*

(W^l) *If $\Gamma, \varphi, \varphi \vdash \Delta$, then $\Gamma, \varphi \vdash \Delta$;*

(W^r) *If $\Gamma \vdash \Delta, \varphi, \varphi$, then $\Gamma \vdash \Delta, \varphi$;*

(K^l) *If $\Gamma \vdash \Delta$, then $\Gamma, \varphi \vdash \Delta$;*

(K^r) *If $\Gamma \vdash \Delta$, then $\Gamma \vdash \Delta, \varphi$;*

(I) *If Γ is non-empty,¹⁶ $\Gamma \vdash \Gamma$.*

¹⁵We state structural properties in a general fashion, using a variable ' \vdash ' taking as values specific consequence relations (such as $\models_{\mathbf{T}^0}$). $\text{fld}(R)$ is the field of R .

¹⁶Note that, given our official stipulations about sequences (definition 4.2.3), the empty sequence just is the empty set \emptyset .

Proof.

- $(C^l), (C^r)$: Immediate from the field invariance of the sequences.
- $(W^l), (W^r)$: Immediate from the fact that $\{\psi : \psi \in \text{ran}(\Gamma, \varphi, \varphi)\} = \{\psi : \psi \in \text{ran}(\Gamma, \varphi)\}$.
- (K^l) : Immediate from the fact that $\{\psi : \psi \in \text{ran}(\Gamma)\} \subseteq \{\psi : \psi \in \text{ran}(\Gamma, \varphi)\}$.
- (K^r) : Immediate from the fact that $\{\psi : \psi \in \text{ran}(\Delta)\} \subseteq \{\psi : \psi \in \text{ran}(\Delta, \varphi)\}$.
- (I) : If Γ is non-empty and, for every $\varphi \in \text{ran}(\Gamma)$, $\text{val}(\varphi) \in D$, then, for some $\varphi \in \text{ran}(\Gamma)$, $\text{val}(\varphi) \in D$, and so *a fortiori* $\text{val}(\varphi) \in T$ (if Γ is empty, then, even though, for every $\varphi \in \text{ran}(\Gamma)$, $\text{val}(\varphi) \in D$, there is no $\varphi \in \text{ran}(\Gamma)$ such that $\text{val}(\varphi) \in T$).

□

The failure of a particular structural property such as transitivity can thus be achieved in full autonomy from the other usual structural properties.¹⁷ It is too seldom noticed that transitivity itself comes in two flavours, left and right. *Left transitivity* concerns, roughly, the legitimacy of chaining a class of inferences together with another inference once some of the conclusions of the class are jointly sufficient to constitute a component of the *premises* (left part) of the latter inference:

(T^l) If, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash \Delta, \varphi$ and $\Lambda, \Theta \vdash \Xi$, then $\Lambda, \Gamma \vdash \Delta, \Xi$.

¹⁷In the few previous attempts to develop a non-transitive consequence relation, this has not always been so. For example, in the non-transitive system mooted in Smiley [1959], pp. 238–43, (W^l) fails.

Right transitivity concerns, roughly, the legitimacy of chaining an inference together with a class of inferences once some of the premises of the class are jointly sufficient to constitute a component of the *conclusions* (right part) of the former inference, and is obtained by dualizing consequence with entailment in (T^l) :

(T^r) If $\Xi \vdash \Lambda, \Theta$ and, for every $\varphi \in \text{ran}(\Theta)$, $\Delta, \varphi \vdash \Gamma$, then $\Delta, \Xi \vdash \Lambda, \Gamma$.

Given (W^l) and (W^r) , (T^l) implies the *cumulative left-transitivity property*:

(CT^l) If, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash \Delta, \varphi$ and $\Gamma, \Theta \vdash \Delta$, then $\Gamma \vdash \Delta$,

and, given (K^l) and (K^r) , it is implied by it. Analogously, given (W^l) and (W^r) , (T^r) implies the *cumulative right-transitivity property*:

(CT^r) If $\Delta \vdash \Gamma, \Theta$ and, for every $\varphi \in \text{ran}(\Theta)$, $\Delta, \varphi \vdash \Gamma$, then $\Delta \vdash \Gamma$,

and, given (K^l) and (K^r) , it is implied by it.

Definition 4.2.10.

- A sequence of argument forms $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \mathbb{F}_\alpha \rangle$ has the *left-transitivity property* in a logic \vdash iff, if, [for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash \Delta, \varphi$ and the forms of these arguments are $\mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots$ (where each \mathbb{F}_β is such that $\beta < \alpha$ and, for every $\gamma < \alpha$, \mathbb{F}_γ is a \mathbb{F}_β)] and $[\Lambda, \Theta \vdash \Xi$ and the form of this argument is $\mathbb{F}_\alpha]$, then $\Lambda, \Gamma \vdash \Delta, \Xi$.
- An argument form \mathbb{F} has the *reflexive left-transitivity property* in a logic \vdash iff every sequence whose domain has a maximum element and whose range is $\{\mathbb{F}\}$ has the left-transitivity property in \vdash .

- A sequence of argument forms $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \rangle$ has the *full 0-left-transitivity property* in a logic \vdash iff, for every argument form \mathbb{F}_α , $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \mathbb{F}_\alpha \rangle$ has the left-transitivity property in \vdash . An argument form \mathbb{F}_α has the *full 1-left-transitivity property* in a logic \vdash iff, for every sequence of argument forms $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \rangle$, $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \mathbb{F}_\alpha \rangle$ has the left-transitivity property in \vdash .
- An argument form \mathbb{F} has the *full left-transitivity property* in a logic \vdash iff it has the full 1-left-transitivity property in \vdash and every sequence whose range is $\{\mathbb{F}\}$ has the full 0-left-transitivity property in \vdash .

Definition 4.2.11.

- A sequence of argument forms $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \rangle$ has the *right-transitivity property* in a logic \vdash iff, if $[\Xi \vdash \Lambda, \Theta$ and the form of this argument is $\mathbb{F}_0]$ and, [for every $\varphi \in \text{ran}(\Theta)$, $\Delta, \varphi \vdash \Gamma$ and the forms of these arguments are $\mathbb{F}_1, \mathbb{F}_2, \mathbb{F}_3 \dots$], then $\Delta, \Xi \vdash \Lambda, \Gamma$.
- An argument form \mathbb{F} has the *reflexive right-transitivity property* in a logic \vdash iff every sequence whose range is $\{\mathbb{F}\}$ has the right-transitivity property in \vdash .
- An argument form \mathbb{F}_0 has the *full 0-right-transitivity property* in a logic \vdash iff, for every sequence of argument forms $\langle \mathbb{F}_1, \mathbb{F}_2, \mathbb{F}_3 \dots \rangle$, $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \rangle$ has the right-transitivity property in \vdash . A sequence of argument forms $\langle \mathbb{F}_1, \mathbb{F}_2, \mathbb{F}_3 \dots \rangle$ has the *full 1-right-transitivity property* in a logic \vdash iff, for every argument form \mathbb{F}_0 , $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \rangle$ has the right-transitivity property in \vdash .
- An argument form \mathbb{F} has the *full right-transitivity property* in a logic \vdash iff it has the full 0-right-transitivity property in \vdash and every sequence whose range is $\{\mathbb{F}\}$ has the full 1-right-transitivity property in \vdash .

Definition 4.2.12. For every tolerant logic \mathbf{T}^* , an argument form with premises Γ and conclusions Δ is *strict* in \mathbf{T}^* iff, for every \mathbf{T}^* -model \mathfrak{M} , for some $\varphi \in \text{ran}(\Gamma)$, for some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\varphi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\psi)$.

Theorem 4.2.3. *For every tolerant logic \mathbf{T}^* , if \mathbb{F} is strict in \mathbf{T}^* , \mathbb{F} has the full left- and right-transitivity properties.*

Proof. Suppose that \mathbb{F} is strict in \mathbf{T}^* and that it has premises of the form \mathbb{G} and conclusions of the form \mathbb{D} .

- *Every sequence whose range is $\{\mathbb{F}\}$ has the full 0-left-transitivity property.* Suppose that, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash \Delta, \varphi$ and $\Lambda, \Theta \vdash \Xi$, and that the form of the premises of the former arguments is \mathbb{G} and the form of their conclusions is \mathbb{D} . Since \mathbb{F} is strict, Δ, φ is non-empty for each such φ . Thus, given a \mathbf{T}^* -model \mathfrak{M} , one of the two cases holds for each such φ :

- (i) For some $\psi \in \text{ran}(\Gamma)$, for some $\chi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\psi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\chi)$;
- (ii) For some $\psi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\varphi)$.

If case (i) holds for some such φ , then it is clear that, if, for every $\psi \in \text{ran}(\Lambda, \Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta, \Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$ (since $D_{\mathfrak{M}} \subseteq T_{\mathfrak{M}}$ and $T_{\mathfrak{M}}$ is an upper set). If case (ii) holds for every such φ , then, if, for every $\psi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, for every such φ , $\text{val}_{\mathfrak{M}}(\varphi) \in D_{\mathfrak{M}}$ as well (since $D_{\mathfrak{M}}$ is an upper set). Therefore, if, for every $\psi \in \text{ran}(\Lambda, \Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, then, for every $\psi \in \text{ran}(\Lambda, \Theta)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, and so, since $\Lambda, \Theta \vdash \Xi$, for some $\psi \in \text{ran}(\Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$. Therefore, generalizing, $\Lambda, \Gamma \vdash \Delta, \Xi$.

- *\mathbb{F} has the full 1-left-transitivity property.* Suppose that, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash \Delta, \varphi$ and $\Lambda, \Theta \vdash \Xi$, and that the form of the premises of the latter argument is \mathbb{G} and the form of its conclusions is \mathbb{D} . Since \mathbb{F} is strict, Ξ is non-empty. Thus, given a \mathbf{T}^* -model \mathfrak{M} , one of the two cases holds:

- (i) For some $\psi \in \text{ran}(\Lambda)$, for some $\chi \in \text{ran}(\Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\chi)$;
- (ii) For some $\varphi \in \text{ran}(\Theta)$, for some $\psi \in \text{ran}(\Xi)$, $\text{val}_{\mathfrak{M}}(\varphi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\psi)$.

If case (i) holds, then it is clear that, if, for every $\psi \in \text{ran}(\Lambda, \Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta, \Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$ (since $D_{\mathfrak{M}} \subseteq T_{\mathfrak{M}}$ and $T_{\mathfrak{M}}$ is an upper set). If case (ii) holds, then, since for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash \Delta, \varphi$, then, for every such φ , if, for every $\psi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, one of the two cases holds:

- (a) $\text{val}_{\mathfrak{M}}(\varphi) \in T_{\mathfrak{M}}$;
- (b) For some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$.

If case (a) holds for every such φ , then, for every $\varphi \in \text{ran}(\Theta)$, $\text{val}_{\mathfrak{M}}(\varphi) \in T_{\mathfrak{M}}$, and so, since case (ii) holds, for some $\psi \in \text{ran}(\Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$ (since $T_{\mathfrak{M}}$ is an upper set). If case (b) holds for some such φ , we trivially have that, for some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$. Therefore, generalizing twice, $\Lambda, \Gamma \vdash \Delta, \Xi$.

- \mathbb{F} has the full 0-right-transitivity property. Suppose that $\Xi \vdash \Lambda, \Theta$ and, for every $\varphi \in \text{ran}(\Theta)$, $\Delta, \varphi \vdash \Gamma$, and that the form of the premises of the former argument is \mathbb{G} and the form of its conclusions is \mathbb{D} . Since \mathbb{F} is strict, Λ, Θ is non-empty. Thus, given a \mathbf{T}^* -model \mathfrak{M} , one of the two cases holds:

- (i) For some $\psi \in \text{ran}(\Xi)$, for some $\chi \in \text{ran}(\Lambda)$, $\text{val}_{\mathfrak{M}}(\psi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\chi)$;
- (ii) For some $\psi \in \text{ran}(\Xi)$, for some $\varphi \in \text{ran}(\Theta)$, $\text{val}_{\mathfrak{M}}(\psi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\varphi)$.

If case (i) holds, then it is clear that, if, for every $\psi \in \text{ran}(\Delta, \Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Lambda, \Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$ (since $D_{\mathfrak{M}} \subseteq T_{\mathfrak{M}}$ and $T_{\mathfrak{M}}$ is an upper set). If case (ii) holds, then, if, for every $\psi \in \text{ran}(\Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, for some $\varphi \in \text{ran}(\Theta)$, $\text{val}_{\mathfrak{M}}(\varphi) \in D_{\mathfrak{M}}$ as well (since $D_{\mathfrak{M}}$ is an upper set). Since $\Delta, \varphi \vdash \Gamma$, if, for every $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$ as well, then, for some $\psi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$. Therefore, $\Delta, \Xi \vdash \Lambda, \Gamma$.

- Every sequence whose range is $\{\mathbb{F}\}$ has the full 1-right-transitivity property. Suppose that $\Xi \vdash \Lambda, \Theta$ and, for every $\varphi \in \text{ran}(\Theta)$, $\Delta, \varphi \vdash \Gamma$, and

that the form of the premises of the latter arguments is \mathbb{G} and the form of their conclusions is \mathbb{D} . Since \mathbb{F} is strict, Γ is non-empty for each such φ . Thus, given a \mathbf{T}^* -model \mathfrak{M} , one of the two cases holds for each such φ :

- (i) For some $\psi \in \text{ran}(\Delta)$, for some $\chi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\chi)$;
- (ii) For some $\psi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\varphi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}}(\psi)$.

If case (i) holds for some such φ , then it is clear that, if, for every $\psi \in \text{ran}(\Delta, \Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Lambda, \Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$ (since $D_{\mathfrak{M}} \subseteq T_{\mathfrak{M}}$ and $T_{\mathfrak{M}}$ is an upper set). If case (ii) holds for every such φ , since $\Xi \vdash \Lambda, \Theta$, it follows that if, for every $\psi \in \text{ran}(\Xi)$, $\text{val}_{\mathfrak{M}}(\psi) \in D_{\mathfrak{M}}$, one of the two cases holds:

- (a) For some $\varphi \in \text{ran}(\Theta)$, $\text{val}_{\mathfrak{M}}(\varphi) \in T_{\mathfrak{M}}$;
- (b) For some $\psi \in \text{ran}(\Lambda)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$.

If case (a) holds, then, since case (ii) holds, for some $\psi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$ (since $T_{\mathfrak{M}}$ is an upper set). If case (b) holds, we trivially have that, for some $\psi \in \text{ran}(\Lambda)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$. Therefore, generalizing twice, $\Delta, \Xi \vdash \Lambda, \Gamma$.

□

Given that $\models_{\mathbf{T}^0}$ satisfies all the other structural properties needed in order for (CT^l) and (CT^r) to entail (T^l) and (T^r) respectively, the counterexamples we will provide to (T^l) and (T^r) will be counterexamples to (CT^l) and (CT^r) as well. A first kind of counterexample to (T^l) and (T^r) will emerge by studying further properties of \mathbf{T}^0 relating to specific operators.

Some elementary rules and laws for the sentential operators are already valid in \mathbf{T}^0 . For \rightarrow , we have from $(\text{impl}_0^{\leftarrow})$:

Theorem 4.2.4. *The rule of modus ponens:*

$(MP_{\rightarrow}) \quad \varphi \rightarrow \psi, \varphi \models_{\mathbf{T}^0} \psi$

is valid.

Proof. Suppose that $\text{val}(\varphi \rightarrow \psi), \text{val}(\varphi) \in D$. Then, by $(\text{impl}_0^{\leftarrow})$, $\text{val}(\psi) \in \text{tol}(\text{val}(\varphi))$, and so, since $\text{val}(\varphi) \in D$, $\text{val}(\psi) \in T$.

□

Crucially, $(\text{impl}_0^{\leftarrow})$, validating (MP_{\rightarrow}) , already suffices to trigger failures of the transitivity properties for \mathbf{T}^0 :

Theorem 4.2.5. $\models_{\mathbf{T}^0}$ satisfies neither (T^l) nor (T^r) .

Proof.

- (T^l) : Given (MP_{\rightarrow}) , $P_0 \rightarrow Q_0, P_0 \models_{\mathbf{T}^0} Q_0$ and $Q_0 \rightarrow R_0, Q_0 \models_{\mathbf{T}^0} R_0$. Setting $\Gamma = P_0 \rightarrow Q_0, P_0$, $\Delta = \emptyset$, $\Theta = Q_0$, $\Lambda = Q_0 \rightarrow R_0$ and $\Xi = R_0$, (T^l) yields that $Q_0 \rightarrow R_0, P_0 \rightarrow Q_0, P_0 \models_{\mathbf{T}^0} R_0$, which is false, since there are \mathbf{T}^0 -models where $\text{val}(P_0) \in D$, $\text{val}(Q_0) \in \text{tol}(\text{val}(P_0))$ (and so $\text{val}(P_0 \rightarrow Q_0) \in D$) but $\notin D$, $\text{val}(R_0) \in \text{tol}(\text{val}(Q_0))$ (and so $\text{val}(Q_0 \rightarrow R_0) \in D$) but $\notin T$.
- (T^r) : Again, given (MP_{\rightarrow}) , $P_0 \rightarrow Q_0, P_0 \models_{\mathbf{T}^0} Q_0$ and $Q_0 \rightarrow R_0, Q_0 \models_{\mathbf{T}^0} R_0$. Setting $\Xi = P_0 \rightarrow Q_0, P_0$, $\Lambda = \emptyset$, $\Theta = Q_0$, $\Delta = Q_0 \rightarrow R_0$ and $\Gamma = R_0$, (T^r) yields that $Q_0 \rightarrow R_0, P_0 \rightarrow Q_0, P_0 \models_{\mathbf{T}^0} R_0$, which is false as explained above.

□

That such a failure emerges with \rightarrow should not be surprising, given the privileged connection of \rightarrow with the consequence relation $\models_{\mathbf{T}^0}$ (connection determined by the role played by tol in the specification of both $(\text{impl}_0^{\leftarrow})$ and $(\models_{\mathbf{T}^0})$).

Turning to the interaction of this operator with \neg , we note that, thanks to $(\text{impl}_0^{\rightarrow})$, $(\text{impl}_0^{\leftarrow})$ and (tol_2) , the novelty of the framework does not interfere with the validity of the intuitionistically acceptable rule of contraposition:

Theorem 4.2.6. *The rule of contraposition:*

$$(\text{CONTR}) \quad \varphi \rightarrow \psi \models_{\mathbf{T}^0} \neg\psi \rightarrow \neg\varphi$$

is valid.

Proof. Suppose that $\varphi \rightarrow \psi \in D$. Then, by $(\text{impl}_0^{\leftarrow})$, $\text{val}(\psi) \in \text{tol}(\text{val}(\varphi))$. Therefore, by (tol_2) , $\text{neg}(\text{val}(\varphi)) \in \text{tol}(\text{neg}(\text{val}(\psi)))$, and so, by $(\text{impl}_0^{\rightarrow})$, $\text{val}(\neg\psi \rightarrow \neg\varphi) \in D$.¹⁸

□

For \wedge and \vee , we have from (D_0) :

Theorem 4.2.7. *The rules of simplification and addition:*

$$(\text{SIMP}_0) \quad \varphi \wedge \psi \models_{\mathbf{T}^0} \varphi;$$

$$(\text{SIMP}_1) \quad \varphi \wedge \psi \models_{\mathbf{T}^0} \psi;$$

$$(\text{ADD}_0) \quad \varphi \models_{\mathbf{T}^0} \varphi \vee \psi;$$

$$(\text{ADD}_1) \quad \psi \models_{\mathbf{T}^0} \varphi \vee \psi$$

are valid and strict.

¹⁸Related to a comment made earlier (fn 9), we can now appreciate why an equivalence operator \leftrightarrow defined in terms of equiv would not be equivalent to the conjunction of conditionals $(\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)$. Consider a model \mathfrak{M} such that $\text{val}_{\mathfrak{M}}(\varphi) = v_0$, $\text{val}_{\mathfrak{M}}(\psi) = v_1$ and $v_0 \in \text{tol}_{\mathfrak{M}}(v_1)$ and $v_1 \in \text{tol}_{\mathfrak{M}}(v_0)$. Then, by $(\text{equiv}_0^{\rightarrow})$, $\text{val}_{\mathfrak{M}}(\varphi \leftrightarrow \psi) = \text{equiv}_{\mathfrak{M}}(v_0, v_1) = v_2 \in D_{\mathfrak{M}}$. Moreover, by $(\text{impl}_0^{\rightarrow})$, $\text{val}_{\mathfrak{M}}(\varphi \rightarrow \psi) = \text{impl}_{\mathfrak{M}}(v_0, v_1) = v_3 \in D_{\mathfrak{M}}$ and $\text{val}_{\mathfrak{M}}(\psi \rightarrow \varphi) = \text{impl}_{\mathfrak{M}}(v_1, v_0) = v_4 \in D_{\mathfrak{M}}$. However, there is no guarantee that $\text{val}_{\mathfrak{M}}((\varphi \rightarrow \psi) \wedge (\psi \rightarrow \varphi)) = \text{glb}(\{v_3, v_4\}) = v_5 \in D_{\mathfrak{M}}$, let alone that $v_5 = v_2$.

Proof. Immediate from (D_0) .

□

Interestingly, we do have from (D_0) the following restricted transitivity properties for \mathbf{T}^0 :

Theorem 4.2.8. *The properties of conjunction in the premises and disjunction in the conclusions:*

(CP₀) *If $\Gamma, \varphi \models_{\mathbf{T}^0} \Delta$, then $\Gamma, \varphi \wedge \psi \models_{\mathbf{T}^0} \Delta$;*

(CP₁) *If $\Gamma, \psi \models_{\mathbf{T}^0} \Delta$ then $\Gamma, \varphi \wedge \psi \models_{\mathbf{T}^0} \Delta$*

(DC₀) *If $\Gamma \models_{\mathbf{T}^0} \Delta, \varphi$ then $\Gamma \models_{\mathbf{T}^0} \Delta, \varphi \vee \psi$*

(DC₁) *If $\Gamma \models_{\mathbf{T}^0} \Delta, \psi$ then $\Gamma \models_{\mathbf{T}^0} \Delta, \varphi \vee \psi$*

hold.

Proof.

- (CP₀): Immediate from (SIMP₀)'s strictness.
- (CP₁): Immediate from (SIMP₁)'s strictness.
- (DC₀): Immediate from (ADD₀)'s strictness.
- (DC₁): Immediate from (ADD₁)'s strictness.

□

Turning to the interaction of these operators with \neg , we note that, thanks to $(\text{neg}_0^{\Rightarrow})$ and $(\text{neg}_1^{\Rightarrow})$, the novelty of the general framework does not interfere with the validity of the intuitionistically acceptable De Morgan rules:

Theorem 4.2.9. *The De Morgan rules:*

$$(DM_0) \quad \varphi \vee \psi \models_{\mathbf{T}^0} \neg(\neg\varphi \wedge \neg\psi);$$

$$(DM_1) \quad \varphi \wedge \psi \models_{\mathbf{T}^0} \neg(\neg\varphi \vee \neg\psi);$$

$$(DM_2) \quad \neg(\varphi \vee \psi) \models_{\mathbf{T}^0} \neg\varphi \wedge \neg\psi$$

are valid and strict.

Proof.

- (DM₀): By (val_∧), $\text{val}(\neg\varphi \wedge \neg\psi) = \text{glb}(\{\text{val}(\neg\varphi), \text{val}(\neg\psi)\})$, and so, by (neg₀[→]), $\text{val}(\neg\neg\varphi) \preceq \text{val}(\neg(\neg\varphi \wedge \neg\psi))$ and $\text{val}(\neg\neg\psi) \preceq \text{val}(\neg(\neg\varphi \wedge \neg\psi))$, wherefore, by (neg₁[→]), $\text{val}(\varphi) \preceq \text{val}(\neg(\neg\varphi \wedge \neg\psi))$ and $\text{val}(\psi) \preceq \text{val}(\neg(\neg\varphi \wedge \neg\psi))$. But, by (val_∨), $\text{val}(\varphi \vee \psi) = \text{lub}(\{\text{val}(\varphi), \text{val}(\psi)\})$, and so $\text{val}(\varphi \vee \psi) \preceq \text{val}(\neg(\neg\varphi \wedge \neg\psi))$.
- (DM₁): By (val_∧), $\text{val}(\varphi \wedge \psi) = \text{glb}(\{\text{val}(\varphi), \text{val}(\psi)\})$, and so, by (neg₀[→]), $\text{val}(\neg\varphi) \preceq \text{val}(\neg(\varphi \wedge \psi))$ and $\text{val}(\neg\psi) \preceq \text{val}(\neg(\varphi \wedge \psi))$. But, by (val_∨), $\text{val}(\neg\varphi \vee \neg\psi) = \text{lub}(\{\text{val}(\neg\varphi), \text{val}(\neg\psi)\})$, and so $\text{val}(\neg\varphi \vee \neg\psi) \preceq \text{val}(\neg(\varphi \wedge \psi))$. By (neg₀[→]), $\text{val}(\neg\neg(\varphi \wedge \psi)) \preceq \text{val}(\neg(\neg\varphi \vee \neg\psi))$, and so, by (neg₁[→]), $\text{val}(\varphi \wedge \psi) \preceq \text{val}(\neg(\neg\varphi \vee \neg\psi))$.
- (DM₂): By (val_∨), $\text{val}(\varphi \vee \psi) = \text{lub}(\{\text{val}(\varphi), \text{val}(\psi)\})$, and so, by (neg₀[→]), $\text{val}(\neg(\varphi \vee \psi)) \preceq \text{val}(\neg\varphi)$ and $\text{val}(\neg(\varphi \vee \psi)) \preceq \text{val}(\neg\psi)$. But, by (val_∧), $\text{val}(\neg\varphi \wedge \neg\psi) = \text{glb}(\{\text{val}(\neg\varphi), \text{val}(\neg\psi)\})$, and so $\text{val}(\neg(\varphi \vee \psi)) \preceq \text{val}(\neg\varphi \wedge \neg\psi)$.

□

Finally, for \neg itself, we have from (neg₁[→]):

Theorem 4.2.10. *The rule of double-negation introduction:*

$$(DNI) \quad \varphi \models_{\mathbf{T}^0} \neg\neg\varphi$$

is valid and strict.

Proof. Immediate from $(\text{neg}_1^{\Rightarrow})$.

□

Given the strictness in \mathbf{T}^0 of (SIMP_0) , (SIMP_1) , (ADD_0) , (ADD_1) , (DM_0) , (DM_1) , (DM_2) and (DNI) , the result of substituting \rightarrow for $\models_{\mathbf{T}^0}$ in them is a logical truth of \mathbf{T}^0 .

We end here our brief survey of the properties of the basic tolerant logic \mathbf{T}^0 . Even though, as has been seen, \mathbf{T}^0 already enshrines the core idea of tolerant logics, it is manifestly too weak a logic for a vague language, failing to satisfy many properties which any such logic may be reasonably expected to have. Hence, in the following we proceed to strengthen the logic in the usual fashion, by adding further and further constraints on its defining structures, while preserving its tolerance.

4.3 Classical Tolerant Logic

4.3.1 Towards the Classical Tolerant Logic \mathbf{CT}^0

As far as vagueness is concerned, my favoured approach, for reasons concerning the philosophy of the logic of a vague language I cannot develop here, is, to speak somewhat loosely, to validate the full fragment of classical logic consistent with the naive theory of vagueness. To achieve that, we place further constraints on \mathbf{T}^0 -structures, thereby characterizing the class of \mathbf{CT}^0 -structures (and, consequently, of \mathbf{CT}^0 -models), \mathbf{CT}^0 being the “classical” tolerant logic generated by these structures. The investigation of the tolerant logics intermediate between \mathbf{T}^0 and \mathbf{CT}^0 will have to wait for another occasion, but our stepwise way of proceeding will be sufficient to give a flavour of the variety of options available in this area.

We start by imposing the following constraints on D and T :

(D_1^2) If $v_0, v_1 \in D$, then $\text{glb}(\{v_0, v_1\}) \in T$

(D is a *tolerance filter*—informally (and plausibly), if two vague pieces of information are very good, their conjunction is still at least good enough);

(D_2^2) For every $v_0, v_1 \in V$, if $\text{lub}(\{v_0, v_1\}) \in D$, then either $v_0 \in T$ or $v_1 \in T$

(D is *tolerantly prime*—informally (and plausibly), the disjunction of two vague pieces of information can be very good only if at least one of them is at least good enough) and:

(D_3^{\Rightarrow}) For every v , if $v \in D$, then $\text{neg}(v) \in V \setminus T$;

(D_3^{\Leftarrow}) For every v , if $\text{neg}(v) \in V \setminus T$, then $v \in D$

(D is a *tolerance ultrafilter*—informally (and plausibly), a vague piece of information is very good iff its negation is not even good enough).¹⁹

We then require *distributivity of finite glbs over finite lubs*:²⁰

$(\text{glb}/\text{lub}_1^2)$ For every $v_0, v_1, v_2 \in V$, $\text{glb}(\{v_0, \text{lub}(\{v_1, v_2\})\}) = \text{lub}(\{\text{glb}(\{v_0, v_1\}), \text{glb}(\{v_0, v_2\})\})$

—informally (and plausibly), the conjunction of a vague piece of information x with the disjunction of two vague pieces of information y and z is exactly as good as the disjunction of the conjunction of x with y and of the conjunction of x with z .

We finally add the following constraints on neg :

¹⁹Note that, together with $(\text{neg}_1^{\Rightarrow})$, (D_3^{\Rightarrow}) and (D_3^{\Leftarrow}) jointly entail that, for every $v \in V$, $v \in T \setminus D$ iff $\text{neg}(v) \in T \setminus D$ —informally (and plausibly), a vague piece of information is good enough but not very good iff its negation is good enough but not very good.

²⁰Equivalent with *distributivity of finite lubs over finite glbs*:

$(\text{lub}/\text{glb}_1^2)$ For every $v_0, v_1, v_2 \in V$, $\text{lub}(\{v_0, \text{glb}(\{v_1, v_2\})\}) = \text{glb}(\{\text{lub}(\{v_0, v_1\}), \text{lub}(\{v_0, v_2\})\})$.

($\text{neg}_0^{\leftarrow}$) For every $v_0, v_1 \in V$, if $\text{neg}(v_1) \preceq \text{neg}(v_0)$, then $v_0 \preceq v_1$;

($\text{neg}_1^{\leftarrow}$) For every $v \in V$, $\text{neg}(\text{neg}(v)) \preceq v$

(of course, each will do given both ($\text{neg}_0^{\rightarrow}$) and ($\text{neg}_1^{\rightarrow}$))—informally (and plausibly), a vague piece of information is at least as good as the negation of its negation.

Definition 4.3.1. With all these further constraints in place on \mathbf{CT}^0 -structures, the consequence relation on pairs of sequences of wffs $\in WWF_{\mathcal{L}^0}$ constituting the classical tolerant logic \mathbf{CT}^0 ($\models_{\mathbf{CT}^0}$) can finally be defined as follows:

($\models_{\mathbf{CT}^0}$) A sequence of wffs Δ is a consequence of a sequence of wffs Γ iff, for every \mathbf{CT}^0 -model \mathfrak{M} , if, for every $\varphi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}}(\varphi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$.

4.3.2 The Strength of \mathbf{CT}^0

We can then reap the harvest of the new semantics. We will do so by going through the newly introduced constraints focussing on their logical import.

From (D_1^2) and (D_2^2) we have:

Theorem 4.3.1. *The rules of adjunction and disjunction:*

(ADJ) $\varphi, \psi \models_{\mathbf{CT}^0} \varphi \wedge \psi$;

(DISJ) $\varphi \vee \psi \models_{\mathbf{CT}^0} \varphi, \psi$

are valid.

Proof.

- (ADJ): Immediate from (D_1^2).

- (DISJ): Immediate from (D_2^2) .

□

(ADJ) and (DISJ) suffice to trigger new interesting failures of (T^l) and (T^r) for \mathbf{CT}^0 :

Theorem 4.3.2. *(ADJ) and (DISJ) suffice to trigger \rightarrow -free failures of (T^l) and (T^r) for \mathbf{CT}^0 .*

Proof.

- (T^l) : Given (ADJ), $P_0, Q_0 \models_{\mathbf{CT}^0} P_0 \wedge Q_0$ and $R_0, P_0 \wedge Q_0 \models_{\mathbf{CT}^0} R_0 \wedge P_0 \wedge Q_0$. Setting $\Gamma = P_0, Q_0$, $\Delta = \emptyset$, $\Theta = P_0 \wedge Q_0$, $\Lambda = R_0$ and $\Xi = R_0 \wedge P_0 \wedge Q_0$, (T^l) yields that $R_0, P_0, Q_0 \models_{\mathbf{CT}^0} R_0 \wedge P_0 \wedge Q_0$, which is false, since there are \mathbf{CT}^0 -models where $\text{val}(P_0), \text{val}(Q_0), \text{val}(R_0) \in D$, but $\text{val}(P_0 \wedge Q_0) \in T \setminus D$ and $\text{val}(R_0 \wedge P_0 \wedge Q_0) \in V \setminus T$.
- (T^r) : Given (DISJ), $P_0 \vee Q_0 \vee R_0 \models_{\mathbf{CT}^0} P_0 \vee Q_0, R_0$ and $P_0 \vee Q_0 \models_{\mathbf{CT}^0} P_0, Q_0, R_0$. Moreover, $R_0 \models_{\mathbf{CT}^0} P_0, Q_0, R_0$. Setting $\Xi = P_0 \vee Q_0 \vee R_0$, $\Lambda = \emptyset$, $\Theta = P_0 \vee Q_0, R_0$, $\Delta = \emptyset$ and $\Gamma = P_0, Q_0, R_0$, (T^r) yields that $P_0 \vee Q_0 \vee R_0 \models_{\mathbf{CT}^0} P_0, Q_0, R_0$, which is false, since there are \mathbf{CT}^0 -models where $\text{val}(P_0 \vee Q_0 \vee R_0) \in D$ and $\text{val}(P_0 \vee Q_0) \in T \setminus D$, but $\text{val}(P_0), \text{val}(Q_0), \text{val}(R_0) \in V \setminus T$.

□

By themselves, of course, (ADJ) and (DISJ) rule out, respectively, *subvaluational* and *supervaluational* approaches. In a transitive framework, these can be abstractly characterized as follows. A specification of a class of admissible models is assumed (usually, for applications to the problem of vagueness, these are just classical models). It is then said that a sequence of wffs Δ is a subvaluational (supervaluational) consequence of a sequence of wffs Γ iff, for every subclass S of admissible models, if, for every $\varphi \in \text{ran}(\Gamma)$, for some

(every) $\mathfrak{M} \in S$, φ has a designated value in \mathfrak{M} , then, for some $\psi \in \text{ran}(\Delta)$, for some (every) $\mathfrak{M} \in S$, ψ has a designated value in \mathfrak{M} . Given this abstract characterization, the generalization to a non-transitive tolerant framework is straightforward. Given a tolerant consequence relation \mathbf{T}^* , we can define subvaluational and supervaluational versions of \mathbf{T}^* by subvaluating and supervaluating on subclasses of \mathbf{T}^* -models as follows:

($\models_{\mathbf{T}^*}^{sb}$) A sequence of wffs Δ is a $\models_{\mathbf{T}^*}^{sb}$ -consequence of a sequence of wffs Γ iff, for every subclass S of \mathbf{T}^* -models, if, for every $\varphi \in \text{ran}(\Gamma)$, for some $\mathfrak{M} \in S$, $\text{val}_{\mathfrak{M}}(\varphi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta)$, for some $\mathfrak{M} \in S$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$;

($\models_{\mathbf{T}^*}^{sp}$) A sequence of wffs Δ is a $\models_{\mathbf{T}^*}^{sp}$ -consequence of a sequence of wffs Γ iff, for every subclass S of \mathbf{T}^* -models, if, for every $\varphi \in \text{ran}(\Gamma)$, for every $\mathfrak{M} \in S$, $\text{val}_{\mathfrak{M}}(\varphi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta)$, for every $\mathfrak{M} \in S$, $\text{val}_{\mathfrak{M}}(\psi) \in T_{\mathfrak{M}}$.

However, in this essay we will not investigate further subvaluational and supervaluational tolerant logics, returning instead to studying the properties of \mathbf{CT}^0 .

Defining a material-implication operator as usual:

Definition 4.3.2. $\varphi \supset \psi := \neg\varphi \vee \psi$,²¹

from (D_2^2) , (D_3^{\Rightarrow}) and (D_3^{\Leftarrow}) we have:

Theorem 4.3.3. *The rule of modus ponens and the deduction theorem:*

(MP $_{\supset}$) $\varphi \supset \psi, \varphi \models_{\mathbf{CT}^0} \psi$;

(DT $_{\supset}$) *If $\Gamma, \varphi \models_{\mathbf{CT}^0} \psi$, then $\Gamma \models_{\mathbf{CT}^0} \varphi \supset \psi$.*

are valid.

²¹The previous conventions about \rightarrow 's binding force apply to \supset as well.

Proof.

- (MP \supset): Suppose that $\text{val}(\varphi \supset \psi), \text{val}(\varphi) \in D$. Then, by (D_2^2) , either $\text{val}(\text{neg}(\varphi)) \in T$ or $\text{val}(\psi) \in T$. But $\text{val}(\varphi) \in D$, and so, by (D_3^{\Rightarrow}) , $\text{val}(\text{neg}(\varphi)) \notin T$. Therefore, $\text{val}(\psi) \in T$.
- (DT \supset): Suppose that $\Gamma, \varphi \models_{\mathbf{CT}^0} \psi$. Either $\text{val}(\varphi) \in D$ or not. In the first case, since $\Gamma, \varphi \models_{\mathbf{CT}^0} \psi$, if, for every $\chi \in \text{ran}(\Gamma)$, $\text{val}(\chi) \in D$, $\text{val}(\psi) \in T$, and so $\text{lub}(\{\text{val}(\text{neg}(\varphi)), \text{val}(\psi)\}) \in T$. In the second case, by $(\text{neg}_1^{\Rightarrow})$, (D_3^{\Rightarrow}) and (D_3^{\Leftarrow}) , $\text{val}(\text{neg}(\varphi)) \in T$, and so $\text{lub}(\{\text{val}(\text{neg}(\varphi)), \text{val}(\psi)\}) \in T$.

□

From (D_3^{\Rightarrow}) and (D_3^{\Leftarrow}) we also have:

Theorem 4.3.4. *The properties of negation in the premises and negation in the conclusions:*

(NP) *If $\Gamma \models_{\mathbf{CT}^0} \Delta, \varphi$, then $\Gamma, \neg\varphi \models_{\mathbf{CT}^0} \Delta$;*

(NC) *If $\Gamma, \varphi \models_{\mathbf{CT}^0} \Delta$, then $\Gamma \models_{\mathbf{CT}^0} \Delta, \neg\varphi$*

hold.

Proof.

- (NP): Suppose that $\Gamma \models_{\mathbf{CT}^0} \Delta, \varphi$. Either $\text{val}(\varphi) \in T$ or not. In the first case, by $(\text{neg}_1^{\Rightarrow})$ and (D_3^{\Rightarrow}) , $\text{val}(\text{neg}(\varphi)) \notin D$, and so it is vacuously true that, if, for every $\psi \in \text{ran}(\Gamma, \neg\varphi)$, $\text{val}(\psi) \in D$, then, for some $\psi \in \text{ran}(\Delta)$, $\text{val}(\psi) \in T$. In the second case, since $\Gamma \models_{\mathbf{CT}^0} \Delta, \varphi$, it is true that, if, for every $\psi \in \text{ran}(\Gamma, \neg\varphi)$, $\text{val}(\psi) \in D$, then, for some $\psi \in \text{ran}(\Delta)$, $\text{val}(\psi) \in T$.

- (NC): Suppose that $\Gamma, \varphi \models_{\mathbf{CT}^0} \Delta$. Either $\text{val}(\varphi) \in D$ or not. In the first case, since $\Gamma, \varphi \models_{\mathbf{CT}^0} \Delta$, it is true that, if, for every $\psi \in \text{ran}(\Gamma)$, $\text{val}(\psi) \in D$, then, for some $\psi \in \text{ran}(\Delta, \neg\varphi)$, $\text{val}(\psi) \in T$. In the second case, by (D_3^{\leftarrow}) , $\text{val}(\text{neg}(\varphi)) \in T$, and so it is true that, if, for every $\psi \in \text{ran}(\Gamma)$, $\text{val}(\psi) \in D$, then, for some $\psi \in \text{ran}(\Delta, \neg\varphi)$, $\text{val}(\psi) \in T$.

□

Theorem 4.3.5. *The law of excluded middle and the attendant property of exhaustion for the consequence relation:*

$$(\text{LEM}) \quad \emptyset \models_{\mathbf{CT}^0} \varphi \vee \neg\varphi;$$

$$(\text{EXH}) \quad \emptyset \models_{\mathbf{CT}^0} \varphi, \neg\varphi$$

are valid.

Proof.

- (LEM): Either $\text{val}(\varphi) \in D$ or not. In the first case, $\text{lub}(\{\text{val}(\varphi), \text{val}(\neg\varphi)\}) \in D$, and so *a fortiori* $\text{lub}(\{\text{val}(\varphi), \text{val}(\neg\varphi)\}) \in T$. In the second case, by $(\text{neg}_1^{\rightarrow})$, (D_3^{\rightarrow}) and (D_3^{\leftarrow}) , $\text{val}(\neg\varphi) \in T$, and so $\text{lub}(\{\text{val}(\varphi), \text{val}(\neg\varphi)\}) \in T$.
- (EXH): Immediate from the proof of (LEM).

□

Theorem 4.3.6. *The law of non-contradiction and the attendant property of explosion for the consequence relation:*

$$(\text{LNC}) \quad \varphi \wedge \neg\varphi \models_{\mathbf{CT}^0} \emptyset;$$

$$(\text{EXP}) \quad \text{If } \varphi, \neg\varphi \models_{\mathbf{CT}^0} \emptyset$$

are valid.

Proof.

- (LNC): Suppose for *reductio* that $\text{val}(\varphi \wedge \neg\varphi) \in D$. Then $\text{val}(\varphi) \in D$ and $\text{val}(\neg\varphi) \in D$. But, if $\text{val}(\varphi) \in D$, by (D_3^{\Rightarrow}) , $\text{val}(\neg\varphi) \notin D$.
- (EXP): Immediate from the proof of (LNC).²²

□

From $(\text{glb}/\text{lub}_1^2)$ we have:

Theorem 4.3.7. *The distributivity rules:*

$$(\text{DISTR}_{\wedge/\vee}) \quad \varphi \wedge (\psi \vee \chi) \models_{\mathbf{CT}^0} (\varphi \wedge \psi) \vee (\varphi \wedge \chi);$$

$$(\text{DISTR}_{\vee/\wedge}) \quad \varphi \vee (\psi \wedge \chi) \models_{\mathbf{CT}^0} (\varphi \vee \psi) \wedge (\varphi \vee \chi)$$

are valid and strict.

Proof.

- $(\text{DISTR}_{\wedge/\vee})$: Immediate from $(\text{glb}/\text{lub}_1^2)$.
- $(\text{DISTR}_{\vee/\wedge})$: Immediate from $(\text{lub}/\text{glb}_1^2)$.

□

From either $(\text{neg}_0^{\Leftarrow})$ or $(\text{neg}_1^{\Leftarrow})$ we have:

Theorem 4.3.8. *The rule of double-negation elimination:*

²²Note that (LEM) and (EXH) ((LNC) and (EXP)) would come apart in the supervaluational (subvaluational) consequence relation $\models_{\mathbf{CT}^0}^{sp} (\models_{\mathbf{CT}^0}^{sb})$ which uses \mathbf{CT}^0 -models as the class of admissible models.

$$(DNE) \quad \neg\neg\varphi \models_{\mathbf{CT}^0} \varphi$$

is valid and strict.

Proof. Immediate from either $(\text{neg}_0^{\leftarrow})$ or $(\text{neg}_1^{\leftarrow})$.

□

Theorem 4.3.9. *The De Morgan rule:*

$$(DM_3) \quad \neg(\varphi \wedge \psi) \models_{\mathbf{CT}^0} \neg\varphi \vee \neg\psi$$

is valid and strict.

Proof. By (val_\vee) , $\text{val}(\neg\varphi \vee \neg\psi) = \text{lub}(\{\text{val}(\neg\varphi), \text{val}(\neg\psi)\})$, and so, by $(\text{neg}_0^{\rightarrow})$, $\text{val}(\neg(\neg\varphi \vee \neg\psi)) \preceq \text{val}(\neg\neg\varphi)$ and $\text{val}(\neg(\neg\varphi \vee \neg\psi)) \preceq \text{val}(\neg\neg\psi)$. By either $(\text{neg}_0^{\leftarrow})$ or $(\text{neg}_1^{\leftarrow})$, $\text{val}(\neg(\neg\varphi \vee \neg\psi)) \preceq \text{val}(\varphi)$ and $\text{val}(\neg(\neg\varphi \vee \neg\psi)) \preceq \text{val}(\psi)$. But, by (val_\wedge) , $\text{val}(\varphi \wedge \psi) = \text{glb}(\{\text{val}(\varphi), \text{val}(\psi)\})$, and so $\text{val}(\neg(\neg\varphi \vee \neg\psi)) \preceq \text{val}(\varphi \wedge \psi)$. By $(\text{neg}_0^{\rightarrow})$, $\text{val}(\neg(\varphi \wedge \psi)) \preceq \text{val}(\neg\neg(\neg\varphi \vee \neg\psi))$, and so, by either $(\text{neg}_0^{\leftarrow})$ or $(\text{neg}_1^{\leftarrow})$, $\text{val}(\neg(\varphi \wedge \psi)) \preceq \text{val}(\neg\varphi \vee \neg\psi)$.

□

Possibly with exception of (NC), (LEM) and (EXH) (whose defence I cannot undertake here), I submit that all the various rules and laws we have been reviewing can very plausibly be taken to represent correct patterns of reasoning with a vague language: competent speakers do usually abide by—and hold others responsible to—they. It is thus crucial to see how, in \mathbf{CT}^0 , the desired restrictions on (T^l) and (T^r) can be achieved while preserving the full validity of these other rules and laws.

4.3.3 The Weakness of \mathbf{CT}^0

I hope that the foregoing is sufficient to show how rich a fragment of classical logic is preserved by \mathbf{CT}^0 . What is not preserved are exactly those

rules and laws of the operators which encode the transitivity of the classical consequence relation (that some such rules and laws exist is evident from the eliminability of the cut rule in a standard sequent calculus for classical logic). These can be made to emerge already in the extensional fragment $\mathbf{CT}_{\rightarrow}^0$ of \mathbf{CT}^0 (that is, the restriction of \mathbf{CT}^0 to the extensional language $\mathcal{L}_{\rightarrow}^0$ such that $WFF_{\mathcal{L}_{\rightarrow}^0} = WFF_{\mathcal{L}^0} \setminus \{\varphi : \rightarrow \text{ occurs in } \varphi\}$). We have already seen (theorem 4.3.2) that $\mathbf{CT}_{\rightarrow}^0$ fails to satisfy (T^l) . Correspondingly:

Theorem 4.3.10. *The property of material implication in the premises:*

(IP_{\supset}) *If $\Gamma \models_{\mathbf{CT}_{\rightarrow}^0} \Delta, \varphi$ and $\Theta, \psi \models_{\mathbf{CT}_{\rightarrow}^0} \Lambda$, then $\Theta, \Gamma, \varphi \supset \psi \models_{\mathbf{CT}_{\rightarrow}^0} \Lambda, \Delta$*

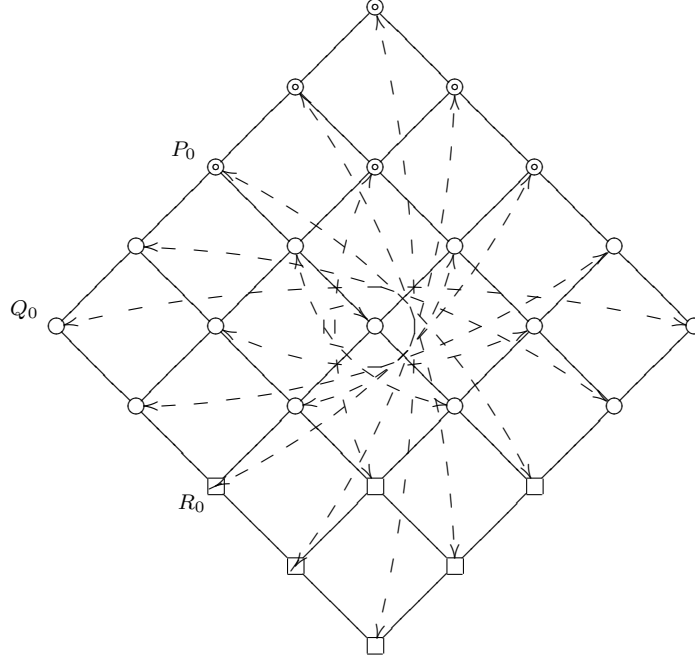
fails.

Proof. We consider the $\mathbf{CT}_{\rightarrow}^0$ -model \mathfrak{M}_0 , where:

- $V_{\mathfrak{M}_0} = \{\langle x, y \rangle : 0 \leq x \leq 4, 0 \leq y \leq 4\}$;
- $D_{\mathfrak{M}_0} = \{\langle x, y \rangle : x + y \geq 6\}$;
- $\preceq_{\mathfrak{M}_0} = \{\langle \pi_0, \pi_1 \rangle : 0\text{co}(\pi_0) \leq 0\text{co}(\pi_1) \text{ and } 1\text{co}(\pi_0) \leq 1\text{co}(\pi_1)\}$ (where $i\text{co}(\pi)$ is the i th coordinate of the ordered pair π);
- $\text{tol}_{\mathfrak{M}_0} = \{\langle \pi_0, \Pi \rangle : \Pi = \{\pi_1 : [0\text{co}(\pi_1) \geq 0\text{co}(\pi_0) - 3 \text{ and } 1\text{co}(\pi_1) \geq 1\text{co}(\pi_0)] \text{ or } [0\text{co}(\pi_1) \geq 0\text{co}(\pi_0) \text{ and } 1\text{co}(\pi_1) \geq 1\text{co}(\pi_0) - 3]\}\}$;
- $O_{\mathfrak{M}_0} = \{\text{neg}_{\mathfrak{M}_0}\}$ (where $\text{neg}_{\mathfrak{M}_0} = \{\langle \pi_0, \pi_1 \rangle : [i\text{co}(\pi_1) = 0 \text{ iff } i\text{co}(\pi_0) = 4] \text{ and } [i\text{co}(\pi_1) = 1 \text{ iff } i\text{co}(\pi_0) = 3] \text{ and } [i\text{co}(\pi_1) = 2 \text{ iff } i\text{co}(\pi_0) = 2] \text{ and } [i\text{co}(\pi_1) = 3 \text{ iff } i\text{co}(\pi_0) = 1]\}$);
- $\text{int}_{\mathfrak{M}_0}(P_0) = \langle 2, 4 \rangle$, $\text{int}_{\mathfrak{M}_0}(Q_0) = \langle 0, 4 \rangle$ and $\text{int}_{\mathfrak{M}_0}(R_0) = \langle 0, 2 \rangle$.

Setting $\Gamma = P_0$, $\Delta = \emptyset$, $\varphi = P_0$, $\Theta = Q_0 \supset R_0$, $\psi = Q_0$ and $\Lambda = R_0$, it's easy to check that the consequent of (IP_{\supset}) is falsified by \mathfrak{M}_0 even though its antecedent holds. \mathfrak{M}_0 may be depicted by the following Hasse diagram

(where double circular nodes indicate the members of $D_{\mathfrak{M}_0}$, simple circular nodes the members of $T_{\mathfrak{M}_0}$, square nodes the members of $V_{\mathfrak{M}_0} \setminus T_{\mathfrak{M}_0}$ and dashed arrows indicate the negation operation):



□

However, we do have the following restricted transitivity property:

Theorem 4.3.11. *(IP $_{\supset}$) holds if the form \mathbb{F}_0 of $\Gamma \models_{\mathbf{CT}_{\rightarrow}^0} \Delta, \varphi$ is such that $\langle \mathbb{F}_0 \rangle$ has the full 0-left-transitivity property and the form \mathbb{F}_1 of $\Theta, \psi \models_{\mathbf{CT}_{\rightarrow}^0} \Lambda$ is such that \mathbb{F}_1 has the full 1-left-transitivity property.*

Proof. Since (MP $_{\supset}$) is valid and $\langle \mathbb{F}_0 \rangle$ has the full 0-left-transitivity property, by (T^l) it follows that $\varphi \supset \psi, \Gamma \models_{\mathbf{CT}_{\rightarrow}^0} \Delta, \psi$. Since \mathbb{F}_1 has the full 1-left-transitivity property, by (T^l), (C^l) and (C^r) it follows that $\Theta, \Gamma, \varphi \supset \psi \models_{\mathbf{CT}_{\rightarrow}^0} \Lambda, \Delta$.

□

We have also seen (theorem 4.3.2) that $\mathbf{CT}_{\rightarrow}^0$ fails to satisfy (T^r). Correspondingly:

Theorem 4.3.12. *The property of disjunction in the premises:*

(DP) *If $\Gamma, \varphi \models_{\mathbf{CT}^0_{\rightarrow}} \Delta$ and $\Theta, \psi \models_{\mathbf{CT}^0_{\rightarrow}} \Lambda$, then $\Theta, \Gamma, \varphi \vee \psi \models_{\mathbf{CT}^0_{\rightarrow}} \Lambda, \Delta$*

fails.

Proof. With \mathfrak{M}_0 as in the proof of theorem 4.3.10, setting $\Gamma = P_0$, $\varphi = \neg P_0$, $\Delta = R_0$, $\Theta = Q_0 \supset R_0$, $\psi = Q_0$ and $\Lambda = R_0$, it's easy to check that the consequent of (DP) fails in \mathfrak{M}_0 even though its antecedent holds. □

However, we do have the following restricted transitivity property:

Theorem 4.3.13. *(DP) holds if the forms $\mathbb{F}_0, \mathbb{F}_1$ of $\Gamma, \varphi \models_{\mathbf{CT}^0_{\rightarrow}} \Delta$ and $\Theta, \psi \models_{\mathbf{CT}^0_{\rightarrow}} \Lambda$ are such that $\langle \mathbb{F}_0, \mathbb{F}_1 \rangle$ has the full 1-right-transitivity property.*

Proof. Since, clearly, full 1-right transitivity is left and right monotonic and commutative, the forms $\mathbb{F}_2, \mathbb{F}_3$ of $\Theta, \Gamma, \varphi \models_{\mathbf{CT}^0_{\rightarrow}} \Lambda, \Delta$ and $\Theta, \Gamma, \psi \models_{\mathbf{CT}^0_{\rightarrow}} \Lambda, \Delta$ are also such that $\langle \mathbb{F}_2, \mathbb{F}_3 \rangle$ has the full 1-right-transitivity property. Since (DISJ) is valid, by (K^l) , (K^r) , (C^l) , (C^r) and (T^r) it follows that $\Theta, \Gamma, \varphi \vee \psi \models_{\mathbf{CT}^0_{\rightarrow}} \Lambda, \Delta$. □

A striking feature of $\mathbf{CT}^0_{\rightarrow}$ (and consequently of \mathbf{CT}^0) is the failure of two traditionally very distinctive restricted transitivity properties:

Theorem 4.3.14. *The properties of suppression of logical truths.²³*

(LTT^l) *If, for every $\varphi \in \text{ran}(\Theta)$, $\emptyset \vdash \Delta, \varphi$ and $\Lambda, \Theta \vdash \Xi$, then $\Lambda \vdash \Delta, \Xi$;*

(LTT^r) *If $\emptyset \vdash \Lambda, \Theta$ and, for every $\varphi \in \text{ran}(\Theta)$, $\Delta, \varphi \vdash \Gamma$, then $\Delta \vdash \Lambda, \Gamma$*

fail.

²³A *logical truth* is simply a logical consequence of the empty sequence.

Proof.

- (LTT^l): We consider the $\mathbf{CT}_{\rightarrow}^0$ -model \mathfrak{M}_1 , which is just as \mathfrak{M}_0 as in proof of theorem 4.3.10, with the exception that, for every $v \in T_{\mathfrak{M}_1} \setminus D_{\mathfrak{M}_1}$, $\text{neg}_{\mathfrak{M}_1}(v) = v$ and possibly with the exception that $\text{int}_{\mathfrak{M}_1}(Q_1) = \langle 4, 0 \rangle$. Setting $\Delta = \emptyset$, $\Theta = Q_0 \vee \neg Q_0, Q_1 \vee \neg Q_1$, $\Lambda = \emptyset$ and $\Xi = (Q_0 \vee \neg Q_0) \wedge (Q_1 \vee \neg Q_1)$, (LTT^l) yields that $\emptyset \models_{\mathbf{CT}^0} (Q_0 \vee \neg Q_0) \wedge (Q_1 \vee \neg Q_1)$, which is false, since $\text{val}_{\mathfrak{M}_1}((Q_0 \vee \neg Q_0) \wedge (Q_1 \vee \neg Q_1)) = \langle 0, 0 \rangle \notin T$.
- (LTT^r): With \mathfrak{M}_1 as above, setting $\Lambda = \emptyset$, $\Theta = Q_0 \vee \neg Q_0$, $\Delta = Q_1 \vee \neg Q_1$ and $\Gamma = (Q_0 \vee \neg Q_0) \wedge (Q_1 \vee \neg Q_1)$, (LTT^r) yields that $\emptyset \models_{\mathbf{CT}^0} (Q_0 \vee \neg Q_0) \wedge (Q_1 \vee \neg Q_1)$, which is false as explained above.

□

I conjecture that the failure of (LTT^l) and (LTT^r) is essentially due to validity of (LEM) (and (EXH)), and so that they hold in weaker tolerant logics. Instead of trying here to establish this claim, I would like to show that, far from indicating a deficiency of the framework developed up to this point, this failure is *unavoidable* in any tolerant logic with certain properties (the following theorem has been inspired by a remark in Field [2007b], p. 2).

Theorem 4.3.15. *For every tolerant logic \mathbf{T}^* , if \mathbf{T}^* validates (LEM), $(\text{DISTR}_{\wedge/\vee})$ (in such a way that $(\text{DISTR}_{\wedge/\vee})$ has the full 1-left-transitivity property and $\langle (\text{DISTR}_{\wedge/\vee}), (\text{DISTR}_{\wedge/\vee}) \rangle$ has the full 1-right-transitivity property) and (LTT^l) (or (LTT^r)), \mathbf{T}^* entails the negation of the conjunction of the characteristic claims of the naive theory of vagueness (1), (2) and (3).*

Proof. From (LEM), we have that either a man with 0 hairs is bald or a man with 0 hairs is not bald and that either a man with 1 hair is bald or a man with 1 hair is not bald, and so, by (LTT^l) (or (LTT^r)), [either a man with 0 hairs is bald or a man with 0 hairs is not bald] and [either a man with 1 hair is bald or a man with 1 hair is not bald]. By the full 1-left-transitivity

property of $(\text{DISTR}_{\wedge/\vee})$, this yields that either $[[\text{either a man with 0 hairs is bald or a man with 0 hairs is not bald}] \text{ and a man with 1 hair is bald}]$ or $[[\text{either a man with 0 hairs is bald or a man with 0 hairs is not bald}] \text{ and a man with 1 hair is bald}]$.

Moreover, by the full 1-right-transitivity property of $\langle(\text{DISTR}_{\wedge/\vee}), (\text{DISTR}_{\wedge/\vee})\rangle$, theorem 4.3.13 yields that $\langle\text{'Either [a man with 0 hairs is bald and a man with 1 hair is bald] or [a man with 0 hairs is not bald and a man with 1 hair is bald]'}, \text{'Either [a man with 0 hairs is bald and a man with 1 hair is bald] or [a man with 0 hairs is not bald and a man with 1 hair is bald]'}\rangle$ follows from $\text{'Either [[either a man with 0 hairs is bald or a man with 0 hairs is not bald] and a man with 1 hair is bald] or [[either a man with 0 hairs is bald or a man with 0 hairs is not bald] and a man with 1 hair is bald]'},$ and so, by the full 1-right-transitivity property of $\langle(\text{ADD}_0), (\text{ADD}_1)\rangle$ and (W^r) , $\text{'Either [either [a man with 0 hairs is bald and a man with 1 hair is bald] or [a man with 0 hairs is not bald and a man with 1 hair is bald]] or [either [a man with 0 hairs is bald and a man with 1 hair is bald] or [a man with 0 hairs is not bald and a man with 1 hair is bald]]'}$ follows from $\text{'Either [[either a man with 0 hairs is bald or a man with 0 hairs is not bald] and a man with 1 hair is bald] or [[either a man with 0 hairs is bald or a man with 0 hairs is not bald] and a man with 1 hair is bald]'}.$ Therefore, by (LTT^l) (or (LTT^r)), $\text{'Either [either [a man with 0 hairs is bald and a man with 1 hair is bald] or [a man with 0 hairs is not bald and a man with 1 hair is bald]] or [either [a man with 0 hairs is bald and a man with 1 hair is bald] or [a man with 0 hairs is not bald and a man with 1 hair is bald]]'}$ is itself a logical truth.

With another 999,999 structurally identical arguments, we reach a (horribly long!) conclusion in disjunctive normal form each of whose disjuncts is of the form $\text{'A man with 0 hairs is/is not bald and a man with 1 hair is/is not bald and a man with 2 hairs is/is not bald... and a man with 1,000,000 hairs is/is not bald'}$. Examples of such disjuncts are: $\text{'A man with 0 hairs is bald and a man with 1 hair is bald and a man with 2 hairs is bald... and$

a man with 1,000,000 hairs is bald’, ‘A man with 0 hairs is bald and a man with 1 hair is bald and a man with 2 hairs is bald. . . and a man with 500,000 hairs is bald and a man with 500,001 hairs is not bald. . . and a man with 1,000,000 hairs is bald’, ‘A man with 0 hairs is bald and a man with 1 hair is bald and a man with 2 hairs is bald. . . and a man with 500,000 hairs is bald and a man with 500,001 hairs is not bald and a man with 500,002 hairs is bald . . . and a man with 1,000,000 hairs is bald’ etc.

Each disjunct represents in effect a complete decision (either in the positive or in the negative), for every i [$i : 0 \leq i \leq 1,000,000$], of the question whether a man with i hairs is bald (we will thus henceforth call any such disjunction ‘a *completeness disjunction*’). Some disjuncts violate (i.e. entail the negation of) the monotonicity principle which—under our current simplifying assumptions—holds for ‘A man with ξ hairs is bald’; all of them violate (3), the tolerance principle for ‘A man with ξ hairs is bald’ to which the naive theory of vagueness is committed,²⁴ save for the “trivialist” ‘A man with 0 hairs is bald and a man with 1 hair is bald and a man with 2 hairs

²⁴Might one not want to reject the implicit reasoning here, and contend that, just as the “intolerant” conclusion that, for some number i [$i : 0 \leq i \leq 999,999$], it is both the case that a man with i hairs is bald and that a man with $i + 1$ hairs is not bald (or *vice versa*) should not follow from the fact that, for every number i [$i : 0 \leq i \leq 1,000,000$], it is either the case that a man with i hairs is bald or it is the case that a man with i hairs is not bald, the equally intolerant conclusion that, for some numeral ι [$\iota : \iota = '0'$ or $\iota = '1'$ or $\iota = '2' \dots$ or $\iota = '999,999'$], it is both the case that ‘A man with ι hairs is bald’ occurs in a disjunct and it is the case that ‘A man with $\iota + 1$ hairs is not bald’ occurs in the disjunct (or *vice versa*) should not follow from the fact that, for every numeral ι [$\iota : \iota = '0'$ or $\iota = '1'$ or $\iota = '2' \dots$ or $\iota = '1,000,000'$], it is either the case that ‘A man with ι hairs is bald’ occurs in a disjunct or it is the case that ‘A man with ι hairs is not bald’ occurs in the disjunct? It is true that the pattern of reasoning presupposed here had better be (and is) generally invalid in a tolerant logic, and it is true that we cannot feasibly secure a guarantee for all such conclusions by directly inspecting each disjunct in question to find out which particular pairs of conjuncts breach tolerance. However, given the classicality of our metalanguage, ‘ ξ_0 occurs in ξ_1 ’ must be assumed to behave classically (and hence, from the point of view of the naive theory, to be precise). The instances of the pattern of reasoning presupposed here are thus unexceptionable.

is bald... and a man with 1,000,000 hairs is bald' and the "nihilist" 'A man with 0 hairs is not bald and a man with 1 hair is not bald and a man with 2 hairs is not bald... and a man with 1,000,000 hairs is not bald', which violate (2) and (1) respectively.

It is actually easy to see that, employing as they do only (SIMP₀) and (SIMP₁), the violations in questions are not only such that every disjunct entails the negation of one or another characteristic claim of the naive theory (and so, by more (SIMP₀), (SIMP₁) and (neg₀[→]), the negation of the conjunction of these claims), but also such that the sequence of the relevant arguments has the 1-right-transitivity property. By theorem 4.3.13, the completeness disjunction itself entails the negation of the conjunction of the characteristic claims of the naive theory.

□

Notice that the antecedent of theorem 4.3.15 is satisfied by **CT**_→⁰ (and, consequently, by **CT**⁰), since (DISTR_{∧/∨}) is validated in **CT**_→⁰ by (glb/lub₁²), which makes ⟨(DISTR_{∧/∨})⟩ strict. Now, consider that there would seem to be no reason why acceptance of a tolerant logic **T**^{*} should preclude one from also accepting the following constraint of *non-acceptance of entailers of contradictions*:

(NEC) If $\Gamma \vdash_{\mathbf{T}^*}$ 'It is not the case that φ ', one ought not to accept φ and accept all the coordinates of Γ .²⁵

However, given (NEC), a theorist who would like to accept the conjunction of all the three characteristic claims of the naive theory ought not to accept the completeness disjunction. Since this is however entailed, together with

²⁵Throughout, I will remain deliberately neutral between wide-scope and narrow-scope readings of a rational requirement flowing from logical consequence (see Broome [1999] on the importance of this distinction). The arguments go through on either disambiguation (see section 5.4.3 for more on the normativity of logical consequence in a non-transitive framework).

the other rules, by (LTT^l) (or (LTT^r)), we can conclude that such principles are not to be had.

4.4 The Consistency of the Naive Theory of Vagueness

4.4.1 Consistency in a Tolerant Framework

Building on what seems to be a plausible model of our use of a vague language, we constructed the basic tolerant logic \mathbf{T}^0 . We then proceeded to develop stronger and stronger systems in order to capture more and more of what seem to be correct patterns of reasoning with a vague language, while at the same time taking care of preserving the hallmark of the weakness of a tolerant logic—namely, the non-transitivity of the consequence relation. The result has been the logic \mathbf{CT}^0 . It is now time to show that \mathbf{CT}^0 is in effect suitable to fulfil the theoretical task for which it has been developed: that of making the naive theory of vagueness consistent once it is assumed as the background logic of the theory.

The informal presentation of the naive theory of vagueness left it rather unspecific what it exactly amounts to. Choosing a particular example, we can be more precise. Keeping in mind the expressive limitations of \mathcal{L}^0 , I propose to consider as a simple paradigmatic example of a zeroth-order naive theory of vagueness the theory \mathcal{N}^0 based on the following axioms (where P_i translates into \mathcal{L}^0 the English ‘ i is a small natural number’ and this is used in a context where 0 is an indisputable positive case and 2 an indisputable negative case):

$$(\mathbf{N}^{0^p}) \ P_0;$$

$$(\mathbf{N}^{0^n}) \ \neg P_2;$$

$$(\mathbf{N}^{0^{t_0}}) \ P_0 \supset P_1;$$

$(N^{0^{t_1}}) \ P_1 \supset P_2.$

\mathcal{N}^0 is palpably a zeroth-order naive theory of vagueness inasmuch as it contains all the three characteristic claims of such a theory (see section 4.1): the existence of positive cases, the existence of negative cases and the non-existence of a sharp boundary between them.

Given the non-standard properties of \models_{CT^0} and the existence of two non-coincident kinds of “good” values in the underlying semantics (the members of D and the members of T), it is not immediate how the intuitive notion of *consistency* might best be captured in the present framework. Indeed, it is easily seen that the framework allows for a multiplicity of different, non-equivalent definitions of the consistency of a sequence. Such definitions will differ in their strength and there is every reason to think that different definitions will prove useful for different theoretical purposes. In this essay, I propose however to focus for simplicity’s sake only on one such definition which appears to be very natural and deeply embedded in our inferential practices. Consider that, given a consequence relation \mathbf{L} , the following constraint linking consequence with *acceptance* and *rejection* should hold:

(AR) If $\Gamma \vdash_{\mathbf{L}} \Delta$, then one ought not to accept all the coordinates of Γ and reject all the coordinates of Δ .²⁶

If Δ is empty, (AR) comes down to the condition that one ought not to accept all the coordinates of Γ , which in turn seems to capture well at least one construal of the intuitive notion of consistency. On such a reading, a sequence Γ is inconsistent in a logic \mathbf{L} iff $\Gamma \vdash_{\mathbf{L}} \emptyset$. Notice that a parallel

²⁶Whenever rejecting φ implies accepting ‘It is not the case that φ ’ and φ entails ‘It is not the case that it is not the case that φ ’, the relevant instance of (NEC) implies the corresponding (single-conclusion) instance of (AR). Conversely, whenever accepting φ implies rejecting ‘It is not the case that φ ’, the relevant (single-conclusion) instance of (AR) implies the corresponding instance of (NEC) (again, see section 5.4.3 for more on the normativity of logical consequence in a non-transitive framework).

argument can be run for the notion of *validity*: if Γ is empty, (AR) comes down to the condition that one ought not to reject all the coordinates of Δ , which in turn seems to capture well at least one construal of the intuitive notion of validity. On such a reading, a sequence Δ is valid in a logic \mathbf{L} iff $\emptyset \vdash_{\mathbf{L}} \Delta$.

Applying these definitions to \mathbf{CT}^0 (and, more generally, to any logic definable in our framework), we obtain that $(\models_{\mathbf{CT}^0})$ reduces the consistency of Γ to every coordinate of Γ being in $D_{\mathfrak{M}}$ for some \mathbf{CT}^0 -model \mathfrak{M} , and that it reduces the validity of Δ to some coordinate of Δ being in $T_{\mathfrak{M}}$ for every \mathbf{CT}^0 -model \mathfrak{M} . Given the properties of $\models_{\mathbf{CT}^0}$, the *theory* based on Γ (that is, the set of (single-conclusion) logical consequences of Γ)²⁷ will then be guaranteed to have all of its members in $T_{\mathfrak{M}}$.

Note in particular that, given $(\models_{\mathbf{CT}^0})$, premises (and conclusions) are not really “put together” when evaluating whether consequence holds or not: what is relevant in such evaluation is only whether, for every premise, the value of the premise belongs to D rather than whether the value resulting from conjunctively “putting together” the premises (the glb of the set of values of the premises) itself belongs to D (analogously, what is relevant in such evaluation is only whether, for some conclusion, the value of the conclusion belongs to T rather than whether the value resulting from disjunctively “putting together” the conclusions (the lub of the set of values of the conclusions) itself belongs to T). This independence seems desirable, as the requirement that the values of the premises (or of the conclusions)

²⁷A more general notion of theory exploiting the multiple-conclusion setting would have the theory of Γ be the set of sequences which are consequences of Γ . I skip over such niceties here. Notice also that, in a non-transitive framework, we cannot sensibly employ the usual, stronger definition of the theory of Γ , which identifies it with the *closure* of Γ under logical consequence (that is, the smallest set of sentences \mathcal{T} such that:

- (i) For every φ , if $\Gamma \vdash \varphi$, then $\varphi \in \mathcal{T}$;
- (ii) For every φ and Θ (such that $\text{ran}(\Theta) = T$), if $\Theta \vdash \varphi$, then $\varphi \in \mathcal{T}$).

I discuss the philosophical and technical implications of this circumstance in section 5.4.6.

undergo any logical operation (such as glb or lub) before being evaluated for consequence would seem to build already into the very definition of the consequence relation a form of transitivity contrary to the spirit of tolerant logics. Therefore, as regards consistency in particular, $(\models_{\mathbf{CT}^0})$ only requires there to be a \mathbf{CT}^0 -model \mathfrak{M} where the value of every premise belongs to $D_{\mathfrak{M}}$ —it does not require there to be a \mathbf{CT}^0 -model \mathfrak{M} where the value of the conjunction of every premise itself belongs to $D_{\mathfrak{M}}$ (which, given the characteristic “lowering” behaviour that the conjunction operation has in \mathbf{CT}^0 , would amount to a more exacting requirement).²⁸ It thus only remains to show that there is indeed such a model.

4.4.2 A Model of Tolerance

For simplicity’s sake, we focus on the consistency result for $\mathbf{CT}_{\rightarrow}^0$. The extension to full \mathbf{CT}^0 is straightforward.

Theorem 4.4.1. *The axiomatic base of $\mathcal{N}^0 \langle (N^{0p}), (N^{0n}), (N^{0t_0}), (N^{0t_1}) \rangle$ is consistent in $\mathbf{CT}_{\rightarrow}^0$.*

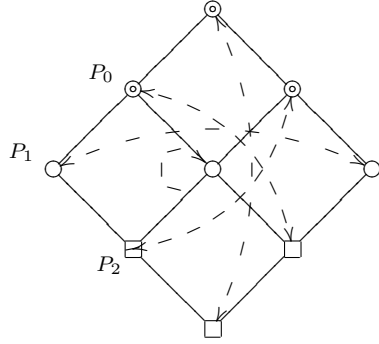
Proof. We consider the $\mathbf{CT}_{\rightarrow}^0$ -model \mathfrak{C}^0 where:

- $V_{\mathfrak{C}^0} = \{\langle x, y \rangle : 0 \leq x \leq 2, 0 \leq y \leq 2\};$
- $D_{\mathfrak{C}^0} = \{\langle x, y \rangle : x + y \geq 3\};$
- $\preceq_{\mathfrak{C}^0} = \{\langle \pi_0, \pi_1 \rangle : 0\text{co}(\pi_0) \leq 0\text{co}(\pi_1) \text{ and } 1\text{co}(\pi_0) \leq 1\text{co}(\pi_1)\};$
- $\text{tol}_{\mathfrak{C}^0} = \{\langle \pi_0, \Pi \rangle : \Pi = \{\pi_1 : [0\text{co}(\pi_1) \geq 0\text{co}(\pi_0) - 1 \text{ and } 1\text{co}(\pi_1) \geq 1\text{co}(\pi_0)] \text{ or } [0\text{co}(\pi_1) \geq 0\text{co}(\pi_0) \text{ and } 1\text{co}(\pi_1) \geq 1\text{co}(\pi_0) - 1]\}\};$
- $O_{\mathfrak{C}^0} = \{\text{neg}_{\mathfrak{C}^0}\}$, where $\text{neg}_{\mathfrak{C}^0} = \{\langle \pi_0, \pi_1 \rangle : [ico(\pi_1) = 0 \text{ iff } ico(\pi_0) = 2] \text{ and } [ico(\pi_1) = 1 \text{ iff } ico(\pi_0) = 1]\};$

²⁸This is not to say of course that such consequence relations lack theoretical interest. Their study must however wait for another occasion.

- $\text{int}_{\mathfrak{C}^0}(P_0) = \langle 1, 2 \rangle$, $\text{int}_{\mathfrak{C}^0}(P_1) = \langle 0, 2 \rangle$ and $\text{int}_{\mathfrak{C}^0}(P_2) = \langle 0, 1 \rangle$.

It's easy to check that \mathfrak{C}^0 is indeed a $\mathbf{CT}_{\rightarrow}^0$ -model for the axiomatic base of \mathcal{N}^0 . \mathfrak{C}^0 may be depicted by the following Hasse diagram (notational conventions as in proof 4.3.10):



□

4.5 Going First-Order

4.5.1 Vagueness and First-Order Expressive Power

In developing a revisionary logic for vagueness, matters can hardly be left at the sentential level. This is *generally* so—no matter what the particular logic developed is—because some problematic claims expressible in a vague language, such as the claim that *there is* a sharp boundary between the bald and the non-bald, can only be adequately regimented in a first-order language, and because the notion of *identity* itself is apparently subject to sorites paradoxes which only employ plausible principles pertaining to the identity predicate, so that a specific treatment of the predicate would appear to be called for. This need is of course at its most acute if the general thrust of the logical revision consists in declaring sorites paradoxes *invalid* rather than *unsound*, for then fault must be found not simply in the alleged truth of at least some members of a set of plausible identity claims, but rather in

the alleged validity of at least some members of the set of the classical rules of inference governing the identity predicate.

An informal presentation of the sorites paradox for identity goes as follows (see Priest [1991]; Priest [1998]). Let us use the functor ‘the baldness status of a man with ξ hairs’ in order to categorize numbers of hairs of hairy and less hairy men in a soritical series according to the men’s baldness, just as the functor ‘the colour of ξ ’ can be used to categorize red and less red patches in a soritical series according to the patches’ colour. Let us also assume for the time being that the categorization is so coarse-grained as to encompass only two categories: the baldness status of baldness, enjoyed by all and only those numbers such that a man with one of those numbers of hairs is bald, and the baldness status of non-baldness, enjoyed by all and only those numbers such that a man with one of those numbers of hairs is not bald.

Consider then the premises:

- (9) The baldness status of a man with 0 hairs is not the same as the baldness status of a man with 1,000,000 hairs;
- (10) The baldness status of a man with i hairs is the same as the baldness status of a man with $i + 1$ hairs.

The naive theory of vagueness is committed to both these premises. To see in particular that it is committed to (10), consider that, given the current coarse-grained understanding of the categorization induced by ‘the baldness status of a man with ξ hairs’, the following *abstraction* principle strikes us as a conceptual truth:

- (ABS⁻) The baldness status of a man with i hairs is the same as the baldness status of a man with j hairs iff, [a man with i hairs is bald iff a man with j hairs is bald].

But, given (ABS⁻), the falsity of (10) would require that a man with i hairs be bald while a man with $i + 1$ hairs be not bald—that is, it would require

the falsity of (3), to which we have already seen that the naive theory is committed.

However, from (10) we have that the baldness status of a man with 0 hairs is the same as the baldness status of a man with 1 hair. From (10) we also have that the baldness status of a man with 1 hair is the same as the baldness status of a man with 2 hairs, which, together with the previous lemma that the baldness status of a man with 0 hairs is the same as the baldness status of a man with 1 hair, yields that the baldness status of a man with 0 hairs is the same as the baldness status of a man with 2 hairs. The rule appealed to here is the rule of *indiscernibility of identicals*, which allows one to infer $\Phi(\tau_1)$ (in our case, ‘The baldness status of a man with 0 hairs is the same as the baldness status of a man with 2 hairs’) from $\Phi(\tau_0)$ (in our case, ‘The baldness status of a man with 1 hair is the same as the baldness status of a man with 2 hairs’) and ‘ τ_0 is the same as τ_1 ’ (in our case, ‘The baldness status of a man with 0 hairs is the same as the baldness status of a man with 1 hair’). With another 999,997 structurally identical arguments, we reach the conclusion that the baldness status of a man with 0 hairs is the same as the baldness status of a man with 999,999 hairs. From (10) we also have that the baldness status of a man with 999,999 hairs is the same as the baldness status of a man with 1,000,000 hairs, which, together with the previous lemma that the baldness status of a man with 0 hairs is the same as the baldness status of a man with 999,999 hairs, yields that the baldness status of a man with 0 hairs is the same as the baldness status of a man with 1,000,000 hairs. It would then seem that the contradictory of (9) follows simply from (10).

At the expenses of some continuity with the foregoing and some not wholly uncontroversial metaphysical assumption, the same point could have been made even without appeal to abstracts introduced by something like (ABS⁻): we could have considered my continuous transformation into a skyscraper, taking a very long finite sequence of subsequent times $t_0, t_1, t_2 \dots t_k$ and calling ‘Elia_{*i*}’ the thing (substance) present at t_i . Then,

for every i, j [$i, j : 0 \leq i \leq j \leq k, j = i + 1$], the naive theory is committed to ‘ Elia_i is the same as Elia_j ’, just as it is committed to the relevant substances’ not ceasing to exist because of nanometrical differences in one atom’s location. But then a structurally identical reasoning could be used to establish the absurd ‘ Elia_0 is the same as Elia_k ’. However, consideration of (ABS^-) does allow us to introduce the further problem of what appear to be sound abstraction principles which abstract on a non-transitive relation, and of what import the truth of these principles would have on the classical properties of identity (in particular, on its transitivity).

There are also reasons *specific* to the particular logical systems developed so far which make their extensions to first-order languages non-trivial. For the logical behaviour of the two traditional first-order quantifiers, universal and particular, can very roughly be characterized as “conjunctive” and “disjunctive” respectively, and we have already had occasion to appreciate the distinctive strength of the conjunction and disjunction operations in tolerant logics, in particular as opposed to the structural operation of “putting together” premises or conclusions expressed by ‘,’ (section 4.4.1). This implies that the simple consistency proof given in theorem 4.4.1 cannot be assumed to go through in all its main aspects also for a suitable first-order formulation of the naive theory of vagueness (having as unique sentence which expresses the non-existence of a sharp boundary between positive and negative cases a sentence saying that *there is* no sharp boundary between positive and negative cases). For that would amount to the consistency of the conjunction of the tolerance conditionals occurring in the axiomatic base of \mathcal{N}^0 —that is to $\text{val}((\mathcal{N}^{0^{t_0}}) \wedge (\mathcal{N}^{0^{t_1}}))$ belonging to $D_{\mathcal{C}^0}$ —and it is easy to see that this is not the case.

4.5.2 Syntax

We extend the zeroth-order language \mathcal{L}^0 to the first-order language \mathcal{L}^1 . In view of the theoretical problems highlighted in section 4.5.1, we proceed

immediately to the study of a first-order language with a designated identity predicate and functors.

Definition 4.5.1. The set $AS_{\mathcal{L}^1}$ of the *atomic symbols* of \mathcal{L}^1 is defined by enumeration as in definition 4.2.1 with the deletion of the first clause and the addition of the following clauses:

- The denumerable set $CONST_{\mathcal{L}^1}$ of individual constants $a_0, a_1, a_2 \dots, b_0, b_1, b_2 \dots, c_0, c_1, c_2 \dots$ is a subset of $AS_{\mathcal{L}^1}$;
- The denumerable set $VAR_{\mathcal{L}^1}$ of individual variables $x_0, x_1, x_2 \dots, y_0, y_1, y_2 \dots, z_0, z_1, z_2 \dots$ is a subset of $AS_{\mathcal{L}^1}$;
- For every i , the denumerable set $FUNCT_{\mathcal{L}^1}^i$ of i ary functors $f_0^i, f_1^i, f_2^i \dots, g_0^i, g_1^i, g_2^i \dots, h_0^i, h_1^i, h_2^i \dots$ is a subset of $AS_{\mathcal{L}^1}$.
 $FUNCT_{\mathcal{L}^1} := \bigcup_{i \in \omega} (FUNCT_{\mathcal{L}^1}^i)$;
- For every i , the denumerable set $PRED_{\mathcal{L}^1}^i$ of i ary predicate constants $P_0^i, P_1^i, P_2^i \dots, Q_0^i, Q_1^i, Q_2^i \dots, R_0^i, R_1^i, R_2^i \dots$ is a subset of $AS_{\mathcal{L}^1}$.
 $PRED_{\mathcal{L}^1} := \bigcup_{i \in \omega} (PRED_{\mathcal{L}^1}^i)$;
- The first-order quantifiers \forall and \exists (universal and particular respectively) belong to $AS_{\mathcal{L}^1}$;
- $,$ belongs to $AS_{\mathcal{L}^1}$.

Note that, given the generality of the framework we will be developing, we cannot assume the definability of either quantifier in terms of the other. We pick a designated 1ary functor in $FUNCT_{\mathcal{L}^1}^1$ (say, f_{100}^1) to serve as successor functor, and denote it with “ $'$ ” (in right-superscript notation). We pick another designated 1ary functor in $FUNCT_{\mathcal{L}^1}^1$ (say, f_{101}^1) to serve as abstraction functor, and denote it with “ $@$ ”. We finally pick a designated 2ary predicate constant in $PRED_{\mathcal{L}^1}^2$ (say, P_{100}^2) to serve as identity predicate, and denote it with “ \approx ” (in infix notation).

Definition 4.5.2. The set $TERM_{\mathcal{L}^1}$ of *terms* of \mathcal{L}^1 can be defined by recursion in the usual way:²⁹

- If $\tau \in CONST_{\mathcal{L}^1}, VAR_{\mathcal{L}^1}$, then $\tau \in TERM_{\mathcal{L}^1}$;
- For every i , if $\tau_0, \tau_1, \tau_2 \dots \tau_{i-1} \in TERM_{\mathcal{L}^1}$ and $\rho^i \in FUNCT_{\mathcal{L}^1}^i$, $\rho^i(\tau_0, \tau_1, \tau_2 \dots \tau_{i-1}) \in TERM_{\mathcal{L}^1}$.

Definition 4.5.3. The set $WFF_{\mathcal{L}^1}$ of *wffs* of \mathcal{L}^1 can be defined by recursion as in definition 4.2.2 with the deletion of the first clause and the addition of the following clauses:³⁰

- For every i , if $\tau_0, \tau_1, \tau_2 \dots \tau_{i-1} \in TERM_{\mathcal{L}^1}$ and $\Phi^i \in PRED_{\mathcal{L}^1}^i$, $\Phi^i(\tau_0, \tau_1, \tau_2 \dots \tau_{i-1}) \in WFF_{\mathcal{L}^1}$.
- If $\varphi \in WFF_{\mathcal{L}^1}$, $\xi \in VAR_{\mathcal{L}^1}$ and neither $\forall \xi$ nor $\exists \xi$ occur in φ , then $(\forall \xi(\varphi)), (\exists \xi(\varphi)) \in WFF_{\mathcal{L}^1}$.
- Nothing else belongs to $WFF_{\mathcal{L}^1}$.

Henceforth, to save on brackets, in addition to the previous conventions I will assume that quantifiers bind as strong as \neg and that right associativity holds for lary operators (including quantifiers) (so that $(\star_0 \star_1 \star_2 \dots \star_i \varphi)$ reads $(\star_0(\star_1(\star_2 \dots \star_i(\varphi) \dots)))$, with each \star_j being a lary operator). I will also drop the brackets of functional and predicative application, the commas of argument composition and the arity-indicating superscripts.

The set $SENT_{\mathcal{L}^1}$ of *sentences* of \mathcal{L}^1 can then be defined as usual:

Definition 4.5.4. $SENT_{\mathcal{L}^1} := WFF_{\mathcal{L}^1} \setminus \{\varphi : \text{for some } \xi, \xi \text{ occurs free in } \varphi\}$.

²⁹Throughout, ‘ τ ’, ‘ σ ’ and ‘ v ’ (possibly with numerical subscripts) are used as metalinguistic variables ranging over $TERM_{\mathcal{L}^1}$; ‘ ρ ’ (possibly with numerical subscripts and superscripts (to indicate arity)) is used as a metalinguistic variable ranging over $FUNCT_{\mathcal{L}^1}$.

³⁰Throughout, ‘ Φ ’ (possibly with numerical subscripts and superscripts (to indicate arity)) is used as metalinguistic variables ranging over $PRED_{\mathcal{L}^1}$; ‘ ξ ’ (possibly with numerical subscripts) is used as a metalinguistic variable ranging over $VAR_{\mathcal{L}^1}$.

4.5.3 Tolerant First-Order Semantic Structures

Definitions 4.2.3, 4.2.4 receive the natural modifications. The basic tolerant semantic structures are now defined as follows.

Definition 4.5.5. A \mathbf{T}^1 -structure \mathfrak{S} for \mathcal{L}^1 is a 6ple $\langle U_{\mathfrak{S}}, V_{\mathfrak{S}}, D_{\mathfrak{S}}, \preceq_{\mathfrak{S}}, \text{tol}_{\mathfrak{S}}, O_{\mathfrak{S}} \rangle$, where $V_{\mathfrak{S}}, D_{\mathfrak{S}}, \preceq_{\mathfrak{S}}, \text{tol}_{\mathfrak{S}}$ are as in definition 4.2.5 and:

- $U_{\mathfrak{S}}$ is a non-empty set of objects (the “universe of discourse”);
- $\preceq_{\mathfrak{S}}$ is as in definition 4.2.5 with $(\text{glb}/\text{lub}_0^2)$ replaced by the stronger condition:

$(\text{glb}/\text{lub}_0^i)$ For every $X \subseteq V_{\mathfrak{S}}$, X has a glb and a lub

$(\preceq_{\mathfrak{S}}$ thus corresponds to a *complete* lattice—this is required in order to guarantee the semantic interpretation of quantified wffs when $U_{\mathfrak{S}}$ is infinite);

- $O_{\mathfrak{S}}$ is as in definition 4.2.5 with the addition that $\text{id}_{\mathfrak{S}} \in O_{\mathfrak{S}}$, where $\text{id}_{\mathfrak{S}} : U_{\mathfrak{S}} \times U_{\mathfrak{S}} \mapsto V_{\mathfrak{S}}$ and:

(id_0) For every $u_0, u_1 \in U_{\mathfrak{S}}$, $\text{id}_{\mathfrak{S}}(u_0, u_1) = \text{id}_{\mathfrak{S}}(u_1, u_0)$.

—informally (and plausibly), the identification of x with y is exactly as good as the identification of y with x . This is arguably a necessary condition on the behaviour of a vague-identity operation such as id —we will add further conditions on it in due course.

A \mathbf{T}^1 -structure can then be used to interpret \mathcal{L}^1 relative to an assignment of values to variables, once it is equipped with an interpretation function for $CONST_{\mathcal{L}^1}$, $FUNCT_{\mathcal{L}^1}$ and $PRED_{\mathcal{L}^1}$ and once suitable recursive clauses are given.

Definition 4.5.6. A \mathbf{T}^1 -model \mathfrak{M} for \mathcal{L}^1 based on a \mathbf{T}^1 -structure \mathfrak{S} is a 7ple $\langle U_{\mathfrak{M}}, V_{\mathfrak{M}}, D_{\mathfrak{M}}, \preceq_{\mathfrak{M}}, \text{tol}_{\mathfrak{M}}, O_{\mathfrak{M}}, \text{int}_{\mathfrak{M}} \rangle$, where $U_{\mathfrak{M}}, V_{\mathfrak{M}}, D_{\mathfrak{M}}, \preceq_{\mathfrak{M}}, \text{tol}_{\mathfrak{M}}$ and $O_{\mathfrak{M}}$ are identical to $U_{\mathfrak{S}}, V_{\mathfrak{S}}, D_{\mathfrak{S}}, \preceq_{\mathfrak{S}}, \text{tol}_{\mathfrak{S}}$ and $O_{\mathfrak{S}}$ respectively, and $\text{int}_{\mathfrak{M}} : \text{CONST}_{\mathcal{L}^1} \cup \text{FUNCT}_{\mathcal{L}^1} \cup \text{PRED}_{\mathcal{L}^1} \mapsto U_{\mathfrak{M}} \cup \bigcup_{i \in \omega} (U_{\mathfrak{M}}^i \times U_{\mathfrak{M}}) \cup \bigcup_{i \in \omega} (U_{\mathfrak{M}}^i \times V_{\mathfrak{M}})$ is an interpretation function for $\text{CONST}_{\mathcal{L}^1}, \text{FUNCT}_{\mathcal{L}^1}$ and $\text{PRED}_{\mathcal{L}^1}$, assigning members of $U_{\mathfrak{M}}$ to members of $\text{CONST}_{\mathcal{L}^1}$, members of $U_{\mathfrak{M}}^i \times U_{\mathfrak{M}}$ to members of $\text{FUNCT}_{\mathcal{L}^1}^i$ and members of $U_{\mathfrak{M}}^i \times V_{\mathfrak{M}}$ to members of $\text{PRED}_{\mathcal{L}^1}^i$.

Definition 4.5.7. An *assignment* is a function $\text{ass} : \text{VAR}_{\mathcal{L}^1} \mapsto U_{\mathfrak{M}}$.

Definition 4.5.8. $\text{ass}_0 \approx_{\xi_0} \text{ass}_1 :=$ for every $\xi_1 \neq \xi_0$, $\langle \xi_1, u \rangle \in \text{ass}_0$ iff $\langle \xi_1, u \rangle \in \text{ass}_1$

Definition 4.5.9. Relative to an assignment ass , $\text{int}_{\mathfrak{M}}$ is extended to a full valuation function $\text{val}_{\mathfrak{M}} : \text{WFF}_{\mathcal{L}^1} \mapsto V_{\mathfrak{M}}$ by the two following recursions:

($\text{val}_{\text{CONST}_{\mathcal{L}^1}}$) If $\tau \in \text{CONST}_{\mathcal{L}^1}$, $\text{val}_{\mathfrak{M}, \text{ass}}(\tau) = \text{int}_{\mathfrak{M}}(\tau)$;

($\text{val}_{\text{VAR}_{\mathcal{L}^1}}$) If $\tau \in \text{VAR}_{\mathcal{L}^1}$, $\text{val}_{\mathfrak{M}, \text{ass}}(\tau) = \text{ass}(\tau)$;

($\text{val}_{\text{FUNCT}_{\mathcal{L}^1}}$) If $\rho \in \text{FUNCT}_{\mathcal{L}^1}$, $\text{val}_{\mathfrak{M}, \text{ass}}(\rho) = \text{int}_{\mathfrak{M}}(\rho)$;

($\text{val}_{\text{TERM}_{\mathcal{L}^1}}$) For every i , if $\tau_0, \tau_1, \tau_2 \dots \tau_{i-1} \in \text{TERM}_{\mathcal{L}^1}$
and $\rho^i \in \text{FUNCT}_{\mathcal{L}^1}^i$, $\text{val}_{\mathfrak{M}, \text{ass}}(\rho^i(\tau_0, \tau_1, \tau_2 \dots \tau_{i-1})) =$
 $\text{val}_{\mathfrak{M}, \text{ass}}(\rho^i)(\text{val}_{\mathfrak{M}, \text{ass}}(\tau_0), \text{val}_{\mathfrak{M}, \text{ass}}(\tau_1), \text{val}_{\mathfrak{M}, \text{ass}}(\tau_2) \dots \text{val}_{\mathfrak{M}, \text{ass}}(\tau_{i-1}));$

($\text{val}_{\text{PRED}_{\mathcal{L}^1}}$) If $\Phi \in \text{PRED}_{\mathcal{L}^1}$, $\text{val}_{\mathfrak{M}, \text{ass}}(\Phi) = \text{int}_{\mathfrak{M}}(\Phi)$;

($\text{val}_{\text{TERMPRED}_{\mathcal{L}^1}}$) For every i , if $\tau_0, \tau_1, \tau_2 \dots \tau_{i-1} \in \text{TERM}_{\mathcal{L}^1}$
and $\Phi^i \in \text{PRED}_{\mathcal{L}^1}^i$, $\text{val}_{\mathfrak{M}, \text{ass}}(\Phi^i(\tau_0, \tau_1, \tau_2 \dots \tau_{i-1})) =$
 $\text{val}_{\mathfrak{M}, \text{ass}}(\Phi^i)(\text{val}_{\mathfrak{M}, \text{ass}}(\tau_0), \text{val}_{\mathfrak{M}, \text{ass}}(\tau_1), \text{val}_{\mathfrak{M}, \text{ass}}(\tau_2) \dots \text{val}_{\mathfrak{M}, \text{ass}}(\tau_{i-1}));$

(val_{\neg}) $\text{val}_{\mathfrak{M}, \text{ass}}(\neg\varphi) = \text{neg}_{\mathfrak{M}}(\text{val}_{\mathfrak{M}, \text{ass}}(\varphi));$

(val_{\wedge}) $\text{val}_{\mathfrak{M}, \text{ass}}(\varphi \wedge \psi) = \text{glb}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}, \text{ass}}(\varphi), \text{val}_{\mathfrak{M}, \text{ass}}(\psi)\});$

$$\begin{aligned}
(\text{val}_\vee) \quad & \text{val}_{\mathfrak{M}, \text{ass}}(\varphi \vee \psi) = \text{lub}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}, \text{ass}}(\varphi), \text{val}_{\mathfrak{M}, \text{ass}}(\psi)\}); \\
(\text{val}_\rightarrow) \quad & \text{val}_{\mathfrak{M}, \text{ass}}(\varphi \rightarrow \psi) = \text{impl}_{\mathfrak{M}}(\text{val}_{\mathfrak{M}, \text{ass}}(\varphi), \text{val}_{\mathfrak{M}, \text{ass}}(\psi)); \\
(\text{val}_\forall) \quad & \text{val}_{\mathfrak{M}, \text{ass}_0}(\forall \xi \varphi) = \text{glb}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\}); \\
(\text{val}_\exists) \quad & \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists \xi \varphi) = \text{lub}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\}); \\
(\text{val}_\simeq) \quad & \text{val}_{\mathfrak{M}, \text{ass}}(\tau \simeq \sigma) = \text{id}_{\mathfrak{M}}(\text{val}_{\mathfrak{M}, \text{ass}}(\tau), \text{val}_{\mathfrak{M}, \text{ass}}(\sigma)).
\end{aligned}$$

4.5.4 First-Order Basic Tolerant Logic

Definition 4.5.10. The consequence relation on pairs of sequences of wffs $\in WFF_{\mathcal{L}^1}$ constituting the first-order basic tolerant logic \mathbf{T}^1 ($\models_{\mathbf{T}^1}$) is defined as follows:

($\models_{\mathbf{T}^1}$) A sequence of wffs Δ is a $\models_{\mathbf{T}^1}$ -consequence of a sequence of wffs Γ ($\Gamma \models_{\mathbf{T}^1} \Delta$) iff, for every \mathbf{T}^1 -model \mathfrak{M} , for every assignment ass , if, for every $\varphi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}, \text{ass}}(\varphi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}, \text{ass}}(\psi) \in T_{\mathfrak{M}}$.

It's easy to check that whatever laws, rules and structural properties hold for \mathbf{T}^0 hold for \mathbf{T}^1 as well. Moreover, for \forall and \exists , we have from (D_0) (letting ' $\varphi[\tau/\sigma]$ ' and its like denote the result of substituting τ for every free occurrence of σ in φ , with the usual restriction that τ be free for σ in φ henceforth being implicitly understood to be in place in order to avoid clashes of bound variables):

Theorem 4.5.1. *The rules of universal instantiation and particular generalization:*

$$(\text{UI}) \quad \forall \xi \varphi \models_{\mathbf{T}^1} \varphi[\tau/\xi]$$

$$(\text{PG}) \quad \varphi \models_{\mathbf{T}^1} \exists \xi \varphi[\xi/\tau]$$

are valid and strict.

Proof. Immediate from (D_0) .

□

Interestingly, we do have from (D_0) the following restricted transitivity properties for \mathbf{T}^1 :

Theorem 4.5.2. *The properties of universal quantifier in the premises and particular quantifier in the conclusions:*

(UP) *If $\Gamma, \varphi \models_{\mathbf{T}^1} \Delta$, then $\Gamma, \forall \xi \varphi[\xi/\tau] \models_{\mathbf{T}^1} \Delta$;*

(PC) *If $\Gamma \models_{\mathbf{T}^1} \Delta, \varphi$ then $\Gamma \models_{\mathbf{T}^1} \Delta, \exists \xi \varphi[\xi/\tau]$*

hold.

Proof.

- (UP): Immediate from (UI)'s strictness.
- (PC): Immediate from (PG)'s strictness.

□

Turning to the interaction of the quantifiers with \neg , we note that, thanks to $(\text{neg}_0^{\rightarrow})$ and $(\text{neg}_1^{\rightarrow})$, the novelty of the general framework does not interfere with the validity of the intuitionistically acceptable quantificational De Morgan rules:

Theorem 4.5.3. *The quantificational De Morgan rules:*

(QDM₀) $\exists \xi \varphi \models_{\mathbf{T}^1} \neg \forall \xi \neg \varphi$;

(QDM₁) $\forall \xi \varphi \models_{\mathbf{T}^1} \neg \exists \xi \neg \varphi$;

$$(\text{QDM}_2) \quad \neg\exists\xi\varphi \models_{\mathbf{T}^1} \forall\xi\neg\varphi$$

are valid and strict.

Proof.

- (QDM₀): By (val_∇), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\neg\varphi) = \text{glb}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\})$, and so, for every such ass_1 , it follows by (neg₀[→]) that $\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\neg\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\neg\varphi)$, wherefore, by (neg₁[→]), $\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\neg\varphi)$ —that is, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\neg\varphi)$ is an upper bound for $\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\}$. But, by (val_∃), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi\varphi) = \text{lub}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\})$, and so $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\neg\varphi)$.
- (QDM₁): By (val_∇), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\varphi) = \text{glb}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\})$, and so, for every such ass_1 , it follows by (neg₀[→]) that $\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\varphi)$ —that is, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\varphi)$ is an upper bound for $\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\}$. But, by (val_∃), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi\neg\varphi) = \text{lub}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\})$, and so $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi\neg\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\forall\xi\varphi)$. By (neg₀[→]), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\neg\forall\xi\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\exists\xi\neg\varphi)$, and so, by (neg₁[→]), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\exists\xi\neg\varphi)$.
- (QDM₂): By (val_∃), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi\varphi) = \text{lub}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\})$, and so, for every such ass_1 , it follows by (neg₀[→]) that $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\exists\xi\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi)$ —that is, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\exists\xi\varphi)$ is a lower bound for $\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\}$. But, by (val_∇), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\neg\varphi) = \text{glb}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\varphi) : \text{ass}_0 \approx_\xi \text{ass}_1\})$, and so $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\exists\xi\varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\neg\varphi)$.

□

Finally, for \approx , we have from (id₀):

Theorem 4.5.4. *The rule of symmetry:*

$$(S^\approx) \quad \tau \approx \sigma \models_{\mathbf{T}^1} \sigma \approx \tau$$

is valid and strict.

Proof. Immediate from (id_0) .

□

Given the strictness in \mathbf{T}^1 of (UI), (PG), (QDM_0) , (QDM_1) , (QDM_2) and (S^\approx) , the result of substituting \rightarrow for $\models_{\mathbf{T}^1}$ in them is a logical truth of \mathbf{T}^1 .

We end here our brief survey of the properties of the first-order basic tolerant logic \mathbf{T}^1 . As in the case of \mathbf{T}^0 , \mathbf{T}^1 is manifestly too weak a logic for a vague language, failing to satisfy many properties which any such logic may be reasonably expected to have. Hence, in the following we proceed to strengthen the logic in the usual fashion, by adding further and further constraints on its defining structures.

4.5.5 First-Order Classical Tolerant Logic

(val_\forall) and (val_\exists) clearly follow the venerable tradition of interpreting universal quantification as an extension of the conjunction operation and particular quantification as an extension of the disjunction operation. The further constraints to be added for characterizing \mathbf{CT}^1 -structures (and, consequently, \mathbf{CT}^1 -models), which define the first-order counterpart \mathbf{CT}^1 of zeroth-order \mathbf{CT}^0 , can thus be expected to mirror and extend those adopted for \mathbf{CT}^0 . Again, the investigation of the quantified tolerant logics intermediate between \mathbf{T}^1 and \mathbf{CT}^1 will have to wait for another occasion, but our stepwise way of proceeding will be sufficient to give a flavour of the variety of options available in this area.

We impose the following constraints on D and T :

$$(D_1^i) \quad \text{If } X \subseteq D, \text{ then } \text{glb}(X) \in T$$

(D is a *tolerance superfilter*—informally (and plausibly), if some vague pieces of information are very good, their conjunction is still at least good enough) and:

(D_2^i) For every $X \subseteq V$, if $\text{lub}(X) \in D$, then, for some $v \in X$, $v \in T$

(D is *tolerantly superprime*—informally (and plausibly), the disjunction of some vague pieces of information can be very good only if at least one of them is at least good enough). Note that (D_1^i) entails (D_1^2) and (D_2^i) entails (D_2^2), and so that these independent constraints of \mathbf{CT}^0 -models can be dispensed with in \mathbf{CT}^1 -models. (D_3^{\rightarrow}), (D_3^{\leftarrow}), ($\text{glb}/\text{lub}_1^2$), ($\text{neg}_0^{\leftarrow}$) and ($\text{neg}_1^{\leftarrow}$) go over as constraints on \mathbf{CT}^1 -models.

In order to get the desired benefits from (D_1^i) and (D_2^i), we must also require that:³¹

(U) For every \mathbf{CT}^1 -model \mathfrak{M} , for every $u \in U_{\mathfrak{M}}$, for some $\delta \in \text{CONST}_{\mathcal{L}^1}$, $\text{int}_{\mathfrak{M}}(\delta) = u$.

Of course, if we want to allow for \mathbf{CT}^1 -models of arbitrary (set) size, the previous definitions concerning $\text{CONST}_{\mathcal{L}^1}$, $\text{TERM}_{\mathcal{L}^1}$ and $\text{WFF}_{\mathcal{L}^1}$ must be modified and reformulated in an adequate theory of classes. I leave such niceties to the interested reader.

We finally add the following constraints on id :

($\text{id}_1^{\rightarrow}$) For every $u_0, u_1 \in U$, if $u_0 = u_1$, then $\text{id}(u_0, u_1) \in D$

—informally (and plausibly), if x and y are completely identical, then the identification of x with y is very good—and:

($\text{id}_2^{\rightarrow}$) For every $u_0, u_1, u_2 \in U$, if $\text{id}(u_0, u_1) \in D$ and $\text{id}(u_1, u_2) \in D$, then $\text{id}(u_0, u_2) \in T$

³¹Throughout, ‘ δ ’ (possibly with numerical subscripts) is used as a metalinguistic variable ranging over $\text{CONST}_{\mathcal{L}^1}$.

—informally (and plausibly), if both the identification of x with y and the identification of y with z are very good, then the identification of x with z is at least good enough.

Definition 4.5.11. With all these further constraints in place on \mathbf{CT}^1 -structures, the consequence relation on pairs of sequences of wffs $\in WWF_{\mathcal{L}}$ constituting the first-order classical tolerant logic \mathbf{CT}^1 ($\models_{\mathbf{CT}^1}$) can finally be defined as follows:

($\models_{\mathbf{CT}^1}$) A sequence of wffs Δ is a consequence of a sequence of wffs Γ iff, for every \mathbf{CT}^1 -model \mathfrak{M} , for every assignment ass , if, for every $\varphi \in \text{ran}(\Gamma)$, $\text{val}_{\mathfrak{M}, \text{ass}}(\varphi) \in D_{\mathfrak{M}}$, then, for some $\psi \in \text{ran}(\Delta)$, $\text{val}_{\mathfrak{M}, \text{ass}}(\psi) \in T_{\mathfrak{M}}$.

We can then reap the harvest of the new semantics. We will do so by going through the newly introduced constraints focussing on their logical import.

From (D_1^i) , (D_2^i) and (U) we have:

Theorem 4.5.5. *The infinitary rules of universal generalization and particular instantiation:*

(UG) $\varphi[\delta_0/\xi], \varphi[\delta_1/\xi], \varphi[\delta_2/\xi] \dots \models_{\mathbf{CT}^0} \forall \xi \varphi;$

(PI) $\exists \xi \varphi \models_{\mathbf{CT}^0} \varphi[\delta_0/\xi], \varphi[\delta_1/\xi], \varphi[\delta_2/\xi] \dots$

(where ‘ $\delta_0, \delta_1, \delta_2 \dots$ ’ is a complete “enumeration” of $\text{CONST}_{\mathcal{L}^1}$ —guaranteed to exist in ZFC—and ξ does not occur bound in φ) are valid.

Proof.

- (UG): Immediate from (D_1^i) and (U) .
- (PI): Immediate from (D_2^i) and (U) .

□

From either $(\text{neg}_0^{\leftarrow})$ or $(\text{neg}_1^{\leftarrow})$ we have:

Theorem 4.5.6. *The quantificational De Morgan rule:*

$$(\text{QDM}_3) \quad \neg \forall \xi \varphi \models_{\mathbf{CT}^1} \exists \xi \neg \varphi$$

is valid and strict.

Proof. By (val_{\exists}) , $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists \xi \neg \varphi) = \text{lub}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg \varphi) : \text{ass}_0 \approx_{\xi} \text{ass}_1\})$, and so, for every such ass_1 , it follows by $(\text{neg}_0^{\rightarrow})$ that $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \exists \xi \neg \varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_1}(\neg \neg \varphi)$. By either $(\text{neg}_0^{\leftarrow})$ or $(\text{neg}_1^{\leftarrow})$, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \exists \xi \neg \varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi)$ —that is, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \exists \xi \neg \varphi)$ is a lower bound for $\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_{\xi} \text{ass}_1\}$. But, by (val_{\forall}) , $\text{val}_{\mathfrak{M}, \text{ass}_0}(\forall \xi \varphi) = \text{glb}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\varphi) : \text{ass}_0 \approx_{\xi} \text{ass}_1\})$, and so $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \exists \xi \neg \varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\forall \xi \varphi)$. By $(\text{neg}_0^{\rightarrow})$, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \forall \xi \varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \neg \exists \xi \neg \varphi)$, and so, by either $(\text{neg}_0^{\leftarrow})$ or $(\text{neg}_1^{\leftarrow})$, $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg \forall \xi \varphi) \preceq \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists \xi \neg \varphi)$.

□

Finally, from $(\text{id}_1^{\rightarrow})$ and $(\text{id}_2^{\rightarrow})$ we have:

Theorem 4.5.7. *The reflexivity and transitivity rules:*

$$(\text{R}^{\rightleftharpoons}) \quad \models_{\mathbf{CT}^1} \tau \rightleftharpoons \tau;$$

$$(\text{T}^{\rightleftharpoons}) \quad \tau \rightleftharpoons \sigma, \sigma \rightleftharpoons v \models_{\mathbf{CT}^1} \tau \rightleftharpoons v$$

are valid.

Proof.

- $(\text{R}^{\rightleftharpoons})$: Immediate from $(\text{id}_1^{\rightarrow})$.
- $(\text{T}^{\rightleftharpoons})$: Immediate from $(\text{id}_2^{\rightarrow})$.

□

4.5.6 Naive Abstraction

Let us use again the functor ‘the baldness status of a man with ξ hairs’ in order to categorize numbers of hairs of hairy and less hairy men in a soritical series according to the men’s baldness. Let us also assume this time that the categorization is so fine-grained as to encompass many categories: the baldness status of 0-baldness (enjoyed by all and only those numbers such that a man with one of those numbers of hairs has roughly the same number of hairs as a man with 0 hairs), the baldness status of 1-baldness (enjoyed by all and only those numbers such that a man with one of those numbers of hairs has roughly the same number of hairs as a man with 1 hair), the baldness status of 2-baldness (enjoyed by all and only those numbers such that a man with one of those numbers of hairs has roughly the same number of hairs as a man with 2 hairs) . . . the baldness status of 1,000,000-baldness (enjoyed by all and only those numbers such that a man with one of those numbers of hairs has roughly the same number of hairs as a man with 1,000,000 hairs). Note that we are still in the dark about the identities of the objects (baldness status) just listed, and so, while the list is certainly complete, it may well be redundant.

We assume that the relation a-man-with- x -hairs-has-roughly-the-same-number-of-hairs-as-a-man-with- y -hairs is an equivalence relation (reflexive, symmetric and transitive):

(REFL^{hairs}) A man with x hairs has roughly the same number of hairs as a man with x hairs;

(SYM^{hairs}) If a man with x hairs has roughly the same number of hairs as a man y hairs, then a man with y hairs has roughly the same number of hairs as a man with x hairs;

(TRANS^{hairs}) If both a man with x hairs has roughly the same number of hairs as a man with y hairs and a man with y hairs has roughly the

same number of hairs as a man with z hairs, then a man with x hairs has roughly the same number of hairs as a man with z hairs.

It is crucial to note that ‘A man with ξ_0 hairs has roughly the same number of hairs as a man with ξ_1 hairs’ is vague, and so a tolerant logic must be used when reasoning with it. Note that, under some plausible additional assumptions, $(\text{TRANS}^{\text{hairs}})$ amounts in effect to a tolerance principle for ‘A man with ξ_0 hairs has roughly the same number of hairs as a man with ξ_1 hairs’. For we can also assume the two following necessary and sufficient conditions for the application of the predicate:

$(\text{NEC}^{\text{hairs}})$ A man with x hairs has roughly the same number of hairs as a man with y hairs only if the absolute value of the difference of x with y is ≤ 10 ;

$(\text{SUFF}^{\text{hairs}})$ A man with x hairs has roughly the same number of hairs as a man with y hairs if the absolute value of the difference of x with y is ≤ 1 .

Consider then the premises:

(11) A man with 0 hairs does not have roughly the same number of hairs as a man with 1,000,000 hairs;

(12) A man with i hairs has roughly the same number of hairs as a man with $i + 1$ hairs.

(11) and (12) are validated by $(\text{NEC}^{\text{hairs}})$ and $(\text{SUFF}^{\text{hairs}})$ respectively.

However, from (12) we have that a man with 0 hairs has roughly the same number of hairs as a man with 1 hair. From (12) we also have that a man with 1 hair has roughly the same number of hairs as a man with 2 hairs, which, together with $(\text{TRANS}^{\text{hairs}})$ and the previous lemma that a man with 0 hairs has roughly the same number of hairs as a man with 1 hair, yields

that a man with 0 hairs has roughly the same number of hairs as a man with 2 hairs. With another 999,997 structurally identical arguments, we reach the conclusion that a man with 0 hairs has roughly the same number of hairs as a man with 999,999 hairs. From (12) we also have that a man with 999,999 hairs has roughly the same number of hairs as a man with 1,000,000 hairs, which, together with the previous lemma that a man with 0 hairs has roughly the same number of hairs as a man with 999,999 hairs, yields that a man with 0 hairs has roughly the same number of hairs as a man with 1,000,000 hairs. It would then seem that the contradictory of (11) follows simply from (12) and (TRANS^{hairs}). Fortunately, it doesn't, as the reasoning just rehearsed implicitly appeals to (T'). Within a tolerant framework, our (very plausible) assumptions about 'A man with ξ_0 hairs has roughly the same number of hairs as a man with ξ_1 hairs' are consistent.

The lesson is that, within a tolerant framework, the *transitivity* of a relation must be sharply distinguished from its *chain transitivity* (to the best of my knowledge, Parikh [1983], p. 247 has been the first to draw this very important distinction). In the classical first-order theory of relations, the transitivity of R :

(TRANS^R) For every x, y, z , if both $x Rs y$ and $y Rs z$, then $x Rs z$.

implies its finite chain transitivity:³²

(CTrans^{R^f}) For every $x_0, x_1, x_2 \dots x_i$, if $x_0 Rs x_1$ and $x_1 Rs x_2$ and $x_2 Rs x_3 \dots$ and $x_{i-1} Rs x_i$, then $x_0 Rs x_i$.

This is not so if the background logic is weakened so as to exhibit suitable failures of transitivity, as is the case in every tolerant logic.

³²An infinitary version of chain transitivity would be:

(CTrans^{Rⁱ}) For every X, x, y , if X is well-ordered by R , x is the minimum element of X under R and y is the lub of X under R , then $x Rs y$.

We do get a hint about the identities of baldness status by considering that, given the current fine-grained understanding of the categorization induced by ‘the baldness status of ξ ’, the following abstraction principle strikes us as a conceptual truth:

(ABS) The baldness status of a man with i hairs is the same as the baldness status of a man with j hairs iff, [a man with i hairs has roughly the same number of hairs as a man with j hairs].

The relation mentioned on the right-hand side is an equivalence relation, and so (ABS) at least avoids the immediate incoherence of abstracting on a relation which is not an equivalence relation. Note however that such an incoherence threatens only on the controversial assumption that identity is itself an equivalence relation—in particular, that it is transitive. Such an assumption holds in any tolerant logic—such as \mathbf{CT}^1 —accepting $(\text{id}_2^{\rightarrow})$ (and so validating (T^{\vartriangleright})), but fails for weaker tolerant logics. We will stick to it given its attractiveness, and show how a first-order naive theory incorporating (ABS) is consistent in \mathbf{CT}^1 (see Shapiro [2006], pp. 165–89 for a stimulating discussion of naive abstraction principles in a transitive framework).

4.5.7 The Consistency and Strength of the First-Order Naive Theory of Vagueness

Benefitting from the expressive resources of \mathcal{L}^1 , I propose to consider as a simple paradigmatic example of a first-order naive theory of vagueness the theory \mathcal{N}^1 based on the following axioms (where $P_0\tau$ translates into \mathcal{L}^1 the English ‘A man with τ hairs is bald’, a_i translates into \mathcal{L}^1 the English ‘ i ’, τ' translates into \mathcal{L}^1 the English ‘the successor of τ ’, $@\tau$ translates into \mathcal{L}^1 the English ‘the baldness status of a man with τ hairs’ and $R_0\tau\sigma$ translates into \mathcal{L}^1 the English ‘A man with τ hairs has roughly the same number of hairs as a man with σ hairs’):

$$(N^{1^p}) \ P_0 a_0;$$

$$(N^{1^n}) \ \neg P_0 a_{1,000,000};$$

$$(N^{1^t}) \ \forall x_0 (P_0 x_0 \supset P_0 x'_0);$$

$$(N^{1^{a^n}}) \ \neg @a_0 \approx @a_{1,000,000};$$

$$(N^{1^{a^t}}) \ \forall x_0 (@x_0 \approx @x'_0);$$

$$(N^{1^{a\Rightarrow}}) \ \forall x_0 \forall x_1 (@x_0 \approx @x_1 \supset R_0 x_0 x_1);$$

$$(N^{1^{a\Leftarrow}}) \ \forall x_0 \forall x_1 (R_0 x_0 x_1 \supset @x_0 \approx @x_1).$$

\mathcal{N}^1 is palpably a first-order naive theory of vagueness inasmuch as it contains all the six characteristic claims of such a theory (sections 4.1, 4.5.1): the existence of positive cases, the existence of negative cases, the non-existence of a sharp boundary between them, the non-identity of the categories of the extreme cases, the identity of the categories of adjacent cases and the necessity and sufficiency of the holding of a non-chain-transitive relation between categorized cases for the identity of the categories they fall under. For simplicity's sake, we focus again on the consistency result for $\mathbf{CT}^1_{\rightarrow}$ (that is, the restriction of \mathbf{CT}^1 to the extensional language $\mathcal{L}^1_{\rightarrow}$ such that $WFF_{\mathcal{L}^1_{\rightarrow}} = WFF_{\mathcal{L}^1} \setminus \{\varphi : \rightarrow \text{ occurs in } \varphi\}$). The extension to full \mathbf{CT}^1 is straightforward.

Theorem 4.5.8. *The axiomatic base of \mathcal{N}^1 $\langle (N^{1^p}), (N^{1^n}), (N^{1^t}), (N^{1^{a^n}}), (N^{1^{a^t}}), (N^{1^{a\Rightarrow}}), (N^{1^{a\Leftarrow}}) \rangle$ is consistent in $\mathbf{CT}^1_{\rightarrow}$.*

Proof. We consider the $\mathbf{CT}^1_{\rightarrow}$ -model \mathfrak{C}_0^1 where:

- $U_{\mathfrak{C}_0^1} = \{a_i : 0 \leq i \leq 1,000,000\} \cup \{b_i : 0 \leq i \leq 1,000,000\};$
- $V_{\mathfrak{C}_0^1} = \text{pow}(\{0, 1, 2\}) \cup \{X : X \in \text{pow}(\{0, 1, 2, 3, 4\}) \text{ and } 0 \in X \text{ and } \text{card}(X) \geq 3 \text{ and, if } \text{card}(X) = 3, \text{ then } 1 \notin X\};$

- $D_{\mathfrak{C}_0^1} = \{\{0, 1, 2, 3, 4\}, \{0, 1, 2\}\};$
- $\preceq_{\mathfrak{C}_0^1} = \{\langle X, Y \rangle : X \subseteq Y\} \setminus \{\langle \{0\}, \{0, 1\} \rangle, \langle \{2\}, \{1, 2\} \rangle\};$
- $\text{tol}_{\mathfrak{C}_0^1} = \{\langle X, Y \rangle : \text{either } Y = \{Z : \text{card}(Z) \geq \text{card}(X) - 1\} \text{ or } [X = \{0, 1, 2\} \text{ and } Y = \{Z : \text{card}(Z) \geq 2\} \cup \{\{0\}, \{2\}\}]\};$
- $O_{\mathfrak{C}_0^1} = \{\text{neg}_{\mathfrak{C}_0^1}, \text{id}_{\mathfrak{C}_0^1}\}$, where $\text{neg}_{\mathfrak{C}_0^1} = \{\langle X, Y \rangle, \langle Y, X \rangle : \text{either } [\text{card}(X) = 5 \text{ and } \text{card}(Y) = 0] \text{ or } [\text{card}(X) = 3 \text{ and } \text{card}(Y) = 1 \text{ and } [\text{either } 0 \in X, Y \text{ or } 1 \in X, Y \text{ or } 2 \in X, Y]] \text{ or } \text{card}(X) = \text{card}(Y) = 2\} \text{ and } \text{id}_{\mathfrak{C}_0^1} \text{ is such that, if } \text{id}_{\mathfrak{C}_0^1}(u_0, u_1) \in D_{\mathfrak{C}_0^1}, \text{ then:}$

- (i) For every $\Phi^i \in \text{PRED}_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1, 2, 3, 4\}$ or $1 \notin \text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$, then $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) [j, k : k - j = i];$
- (ii) For every $\Phi^i \in \text{PRED}_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1, 2\}$, then either $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{0, 1\} [j, k : k - j = i];$
- (iii) For every $\Phi^i \in \text{PRED}_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1\}$, then either $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{1\} [j, k : k - j = i];$
- (iv) For every $\Phi^i \in \text{PRED}_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{1, 2\}$, then either $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{0, 1, 2\} [j, k : k - j = i];$

- (v) For every $\Phi^i \in PRED_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{1\}$, then either $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_0^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{1, 2\} [j, k : k - j = i]$.

Moreover, $\text{id}_{\mathfrak{C}_0^1}$ is such that:

- (i') $\text{id}_{\mathfrak{C}_0^1}(u_0, u_1) = \{0, 1, 2\}$ iff, either $u_0 = u_1$ or, for some $i, j [i : 0 \leq i, j \leq 1,000,000]$, $u_0 = b_i$, $u_1 = b_j$ and $|i - j| \leq 1$;
- (ii') $\text{id}_{\mathfrak{C}_0^1}(u_0, u_1) = \{0, 1\}$ iff, for some $i, j [i : 0 \leq i, j \leq 1,000,000]$, $u_0 = b_i$, $u_1 = b_j$ and $|i - j| = 2$;
- (iii') Otherwise, $\text{id}_{\mathfrak{C}_0^1}(u_0, u_1) = \{1\}$.

It's easy to check that these additional conditions on $\text{id}_{\mathfrak{C}_0^1}$ are consistent with conditions (i)-(v).

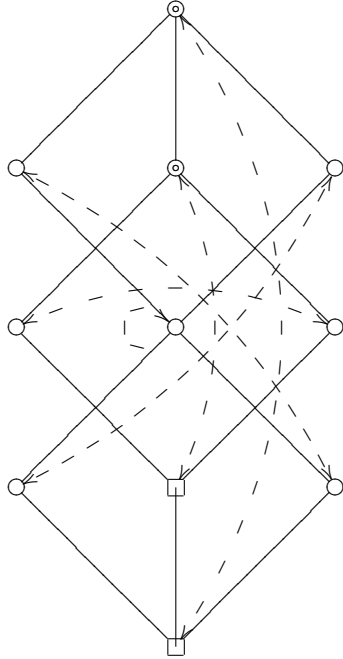
- $\text{int}_{\mathfrak{C}_0^1}$ is such that:

- (i) If, for some $i [i : 0 \leq i \leq 1,000,000]$, $\delta = a_i$ or $\delta = b_i$, then $\text{int}_{\mathfrak{C}_0^1}(\delta) = \delta$;
- (ii) $\text{int}_{\mathfrak{C}_0^1}(@) = \{\langle \langle u_0 \rangle, u_1 \rangle : \text{for some } i [i : 0 \leq i \leq 1,000,000], \text{ either } [u_0 = a_i \text{ and } u_1 = b_i] \text{ or } [u_0 = b_i \text{ and } u_1 = a_i]\}$;
- (iii) $\text{int}_{\mathfrak{C}_0^1}(') = \{\langle \langle u_0 \rangle, u_1 \rangle : \text{for some } i [i : 0 \leq i < 1,000,000], \text{ either } [u_0 = a_i \text{ and } u_1 = a_{i+1}] \text{ or } [u_0 = a_{1,000,000} \text{ and } u_1 = a_{1,000,000}] \text{ or } [u_0 = b_i \text{ and } u_1 = b_{i+1}] \text{ or } [u_0 = b_{1,000,000} \text{ and } u_1 = b_{1,000,000}]\}$;
- (iv) $\text{int}_{\mathfrak{C}_0^1}(P_0) = \{\langle \langle u \rangle, v \rangle : \text{either, [for some } i [i : 0 \leq i \leq 499,997], u = a_i \text{ and } v = \{0, 1, 2, 3, 4\}] \text{ or } [u = a_{499,998} \text{ and } v = \{0, 1, 2\}] \text{ or } [u = a_{499,999} \text{ and } v = \{0, 2, 3\}] \text{ or } [u = a_{500,000} \text{ and } v = \{0, 1\}] \text{ or } [u = a_{500,001} \text{ and } v = \{0\}] \text{ or } [u = a_{500,002} \text{ and } v = \{1\}] \text{ or [for some } i [i : 500,003 \leq i \leq 1,000,000], u = a_i \text{ and } v = \emptyset]; \text{ otherwise, } v = \emptyset\}$;

(v) $\text{int}_{\mathfrak{C}_0^1}(R_0) = \{\langle \langle u_0, u_1 \rangle, v \rangle : \text{for some } i, j \ [i, j : 0 \leq i, j \leq 1,000,000], u_0 = a_i \text{ and } u_1 = a_j \text{ and } [\text{either } |i - j| \leq 1 \text{ and } v = \{0, 1, 2\}] \text{ or } |i - j| = 2 \text{ and } v = \{0, 1\}] \text{ or } |i - j| \geq 2 \text{ and } v = \{1\}]\}; \text{ otherwise, } v = \{1\}\}.$

It's easy to check that these additional conditions on P_0 and R_0 are consistent with conditions (i)-(v) on $\text{id}_{\mathfrak{C}_0^1}$.

It's easy to check that \mathfrak{C}_0^1 is indeed a $\mathbf{CT}_{\rightarrow}^1$ -model for the axiomatic base of \mathcal{N}^1 . \mathfrak{C}_0^1 may be depicted by the following Hasse diagram (notational conventions as in proof 4.3.10):



□

Indeed, not only does a model such as \mathfrak{C}_0^1 provide a proof of the consistency of the axiomatic base of \mathcal{N}^1 in $\mathbf{CT}_{\rightarrow}^1$, but it also validates the other traditional fundamental rule governing identity which had so far escaped our semantic machinery: the rule of indiscernibility of identicals.

Definition 4.5.12. A \mathbf{CT}^1 -model \mathfrak{M} is *quasi- \mathfrak{C}_0^1 -equivalent* iff there exists

an isomorphism between the structure of \mathfrak{M} and the structure of \mathfrak{C}_0^1 possibly with the exception of conditions (i')-(iii') on $\text{id}_{\mathfrak{C}_0^1}$.

Theorem 4.5.9. *The rule of indiscernibility of identicals:*

$$(II) \quad \tau \approx \sigma, \varphi \models_{\mathbf{CT}^1} \varphi[\tau/\sigma]$$

is valid in every quasi- \mathfrak{C}_0^1 -equivalent \mathbf{CT}^1 -model.

Proof. An easy, even though tedious, induction on the complexity of wffs for conditions (i)-(v) on id .

□

(II) is licensed by the conditions (i)-(v) on id , which in effect amount to requiring, informally (and plausibly), that, if both the identification of x with y and the information that y is F are very good, then the information that x is F is at least good enough.³³ Given (II)'s eminent plausibility, I propose that we single out for special attention the strengthening \mathbf{CIIT}^1 of \mathbf{CT}^1 characterized by quasi- \mathfrak{C}_0^1 -equivalent \mathbf{CT}^1 -models.

Other features of \mathfrak{C}_0^1 are also worth remarking upon. Not only does \mathfrak{C}_0^1 validate $(N^{1a\Rightarrow})$ and $(N^{1a\Leftarrow})$, it also validates the strictness of the argument from $@\tau_0 \approx @\tau_1$ to $R_0\tau_0\tau_1$ and *vice versa*. This can easily be seen as requiring that [the relation denoted by R_0 is transitive iff the relation denoted by \approx is transitive], and in \mathfrak{C}_0^1 they are in effect both transitive (even though, of course, not chain-transitive).

Furthermore, \mathfrak{C}_0^1 is such that, unlike many other \mathbf{CT}^1 -models, for every $X \subseteq D_{\mathfrak{C}_0^1}$, $\text{glb}(X) \in D_{\mathfrak{C}_0^1}$. This implies that every quasi- \mathfrak{C}_0^1 -equivalent \mathbf{CT}^1 -model of $\Gamma = \langle \varphi_0, \varphi_1, \varphi_2 \dots \varphi_i \rangle$ is also a model of $\varphi_0 \wedge \varphi_1 \wedge \varphi_2 \dots \wedge \varphi_i$, so that the consistency in \mathbf{CIIT}^1 of the (finite) axiomatic base of a theory \mathcal{T} implies the consistency in \mathbf{CIIT}^1 of the conjunction of the axioms of \mathcal{T} (and the consistency in \mathbf{CIIT}^1 of the axiomatic base $\langle \varphi[\tau_0/\xi], \varphi[\tau_1/\xi], \varphi[\tau_2/\xi] \dots \rangle$

³³Of course, (T^\approx) is just a special case of (II), with $\varphi = \sigma \approx v$.

(where ξ does not occur bound in φ) of a theory \mathcal{T} implies the consistency in **CIIT**¹ of $\forall\xi\varphi$). Dually, \mathfrak{C}_0^1 is such that, unlike many other **CT**¹-models, for every $X \subseteq V_{\mathfrak{C}_0^1} \setminus T_{\mathfrak{C}_0^1}$, $\text{lub}(X) \in V_{\mathfrak{C}_0^1} \setminus T_{\mathfrak{C}_0^1}$. This implies that every quasi- \mathfrak{C}_0^1 -equivalent **CT**¹-model which does not tolerate $\Gamma = \langle \varphi_0, \varphi_1, \varphi_2 \dots \varphi_i \rangle$ (i.e. is such as not to assign a tolerated value to any of $\varphi_0, \varphi_1, \varphi_2 \dots \varphi_i$) does not tolerate $\varphi_0 \vee \varphi_1 \vee \varphi_2 \dots \vee \varphi_i$ either, so that the non-toleration in **CIIT**¹ of the (finite) axiomatic base of a theory \mathcal{T} implies the non-toleration in **CIIT**¹ of the disjunction of the axioms of \mathcal{T} (and the non-toleration in **CIIT**¹ of the axiomatic base $\langle \varphi[\tau_0/\xi], \varphi[\tau_1/\xi], \varphi[\tau_2/\xi] \dots \rangle$ (where ξ does not occur bound in φ) of a theory \mathcal{T} implies the non-toleration in **CIIT**¹ of $\exists\xi\varphi$).

Finally, we can state and prove for **CIIT**¹ a fairly general kind of theorem concerning tolerant logics which reveals the peculiar *incompleteness* of the naive theory of vagueness that they fit so well as to make it consistent. This feature is closely related to that discussed in connection with theorem 4.3.15 (and, in what follows, with theorem 4.5.11). We will study in depth in chapter 6 the philosophical grounds and consequences of such incompleteness—here, we limit ourselves to some of its logical manifestations.

Definition 4.5.13. A logic **L** has the *Lindenbaum property* iff, for every Γ , if $\Gamma \not\vdash_{\mathbf{L}} \emptyset$, then, for some Δ such that, for every φ , if $\varphi \in \text{ran}(\Gamma)$, then $\varphi \in \text{ran}(\Delta)$, and, for every φ , either φ or $\neg\varphi \in \text{ran}(\Delta)$, $\Delta \not\vdash_{\mathbf{L}} \emptyset$.

Theorem 4.5.10. **CIIT**¹ does not have the *Lindenbaum property*.

Proof. Consider $\Gamma = \langle (N^{1^p}), (N^{1^n}), (N^{1^t}), (N^{1^{a^n}}), (N^{1^{a^t}}), (N^{1^{a^{\Rightarrow}}}), (N^{1^{a^{\Leftarrow}}}) \rangle$. As proved in theorem 4.5.8, $\Gamma \not\vdash_{\mathbf{CIIT}^1} \emptyset$. However, for no Δ such that, for every φ , if $\varphi \in \text{ran}(\Gamma)$, then $\varphi \in \text{ran}(\Delta)$, and, for every i [$i : 0 \leq i \leq 1,000,000$], either Pa_i or $\neg Pa_i \in \text{ran}(\Delta)$, $\Delta \not\vdash_{\mathbf{CIIT}^1} \emptyset$. For, since $(N^{1^p}), (N^{1^n}) \in \text{ran}(\Gamma)$, it would follow by familiar reasoning (in the classical metalanguage) that, for some i , $Pa_i \in \text{ran}(\Delta)$ and $\neg Pa_{i+1} \in \text{ran}(\Delta)$. But, since $(N^{1^t}) \in \text{ran}(\Gamma)$, this would be impossible, as $\varphi, \psi, \neg\exists\xi(\varphi \wedge \psi) \vdash_{\mathbf{CIIT}^1} \emptyset$ ((ADJ) has the full 0-left-transitivity property in **CIIT**¹).

□

4.5.8 The Price of Tolerance

A less desirable feature of \mathfrak{C}_0^1 is that $\text{val}_{\mathfrak{C}_0^1, \text{ass}_0}(\exists\xi(P_0\xi \wedge \neg P_0\xi)) = \{0, 1, 2, 3, 4\} \in D_{\mathfrak{C}_0^1}$, since $\text{val}_{\mathfrak{C}_0^1, \text{ass}_1}(P_0\xi \wedge \neg P_0\xi) = \{2\}$ if $\text{ass}_1(\xi) = a_{499,999}$, $\text{val}_{\mathfrak{C}_0^1, \text{ass}_2}(P_0\xi \wedge \neg P_0\xi) = \{1\}$ if $\text{ass}_2(\xi) = a_{499,998}$, $\text{val}_{\mathfrak{C}_0^1, \text{ass}_3}(P_0\xi \wedge \neg P_0\xi) = \{0\}$ if $\text{ass}_3(\xi) = a_{500,001}$, and $\text{lub}(\{\text{val}_{\mathfrak{C}_0^1, \text{ass}_1}(P_0\xi \wedge \neg P_0\xi), \text{val}_{\mathfrak{C}_0^1, \text{ass}_2}(P_0\xi \wedge \neg P_0\xi), \text{val}_{\mathfrak{C}_0^1, \text{ass}_3}(P_0\xi \wedge \neg P_0\xi)\}) = \{0, 1, 2, 3, 4\} \in D_{\mathfrak{C}_0^1}$. It is of the utmost importance to see why this unwelcome result is not simply an accidental feature of \mathfrak{C}_0^1 , but is in a sense *unavoidable* under the current constraints on \mathbf{CT}^1 -models for reasons similar to those exposed in theorem 4.3.15.

Theorem 4.5.11. *No \mathbf{CT}^1 -model is a model both of the conjunction of the coordinates of the axiomatic base of \mathcal{N}^1 and of $\neg\exists\xi(P_0\xi \wedge \neg P_0\xi)$.*

Proof. Suppose that, for some \mathbf{CT}^1 -model \mathfrak{M} and assignment ass_0 , $\text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi(P_0\xi \wedge \neg P_0\xi)) \in V_{\mathfrak{M}} \setminus T_{\mathfrak{M}}$. Then, for every $\text{ass}_1 \approx_{\xi} \text{ass}_0$, $\text{val}_{\mathfrak{M}, \text{ass}_1}(P_0\xi \wedge \neg P_0\xi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi(P_0\xi \wedge \neg P_0\xi))$, and so, by the strictness of (DNE), $\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg P_0\xi \wedge \neg\neg P_0\xi) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi(P_0\xi \wedge \neg P_0\xi))$. Hence, by the strictness of (DM₂), $\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg(P_0\xi \vee \neg P_0\xi)) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi(P_0\xi \wedge \neg P_0\xi))$, and so $\text{lub}_{\mathfrak{M}}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg(P_0\xi \vee \neg P_0\xi)) : \text{ass}_1 \approx_{\xi} \text{ass}_0\}) = \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi\neg(P_0\xi \vee \neg P_0\xi)) \preceq_{\mathfrak{M}} \text{val}_{\mathfrak{M}, \text{ass}_0}(\exists\xi(P_0\xi \wedge \neg P_0\xi)) \in V_{\mathfrak{M}} \setminus T_{\mathfrak{M}}$. Therefore, by (D₃⁼), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\neg\exists\xi\neg(P_0\xi \vee \neg P_0\xi)) \in D_{\mathfrak{M}}$, and so, by the strictness of (QDM₂), $\text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\neg\neg(P_0\xi \vee \neg P_0\xi)) \in D_{\mathfrak{M}}$. Then, for every $\text{ass}_1 \approx_{\xi} \text{ass}_0$, $\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\neg(P_0\xi \vee \neg P_0\xi)) \in D_{\mathfrak{M}}$. Hence, by the strictness of (DNE), $\text{val}_{\mathfrak{M}, \text{ass}_1}(P_0\xi \vee \neg P_0\xi) \in D_{\mathfrak{M}}$ and, since $\text{glb}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(\neg\neg(P_0\xi \vee \neg P_0\xi)) : \text{ass}_1 \approx_{\xi} \text{ass}_0\}) = \text{val}_{\mathfrak{M}, \text{ass}_0}(\forall\xi\neg\neg(P_0\xi \vee \neg P_0\xi)) \in D_{\mathfrak{M}}$, $\text{glb}(\{\text{val}_{\mathfrak{M}, \text{ass}_1}(P_0\xi \vee \neg P_0\xi) : \text{ass}_1 \approx_{\xi} \text{ass}_0\}) \in D_{\mathfrak{M}}$ as well.

Given the strictness of (DISTR _{\wedge/\vee}), a reasoning similar to that exploited in theorem 4.3.15 will now ensure that the corresponding completeness disjunction φ is such that $\text{val}_{\mathfrak{M}, \text{ass}_0}(\varphi) \in D_{\mathfrak{M}}$. By (SIMP₀), (SIMP₁) and (PG) each disjunct entails the negation of one or another coordinate of the axiomatic base of \mathcal{N}^1 (and so, by more (SIMP₀), (SIMP₁) and (neg₀⁼), the

negation of the conjunction ψ of these coordinates). Since the sequence of these arguments has the full 1-right-transitivity property, it follows, by theorem 4.3.13, that φ itself entails $\neg\psi$, and so that $\neg\psi \in T_{\mathfrak{M}}$. But then, by (D_3^{\Rightarrow}) , ψ cannot belong to $D_{\mathfrak{M}}$, and so \mathfrak{M} is not after all a model of ψ .

□

The upshot is that every \mathbf{CT}^1 -model \mathfrak{M} and assignment ass such that $\text{val}_{\mathfrak{M},\text{ass}}(\exists\xi(P_0\xi \wedge \neg P_0\xi)) \in V_{\mathfrak{M}} \setminus T_{\mathfrak{M}}$ do not model ψ (i.e. $\text{val}_{\mathfrak{M},\text{ass}}(\psi) \notin D_{\mathfrak{M}}$); contraposing, every \mathbf{CT}^1 -model \mathfrak{M} and assignment ass which do model ψ will also tolerate $\exists\xi(P_0\xi \wedge \neg P_0\xi)$ (i.e. $\text{val}_{\mathfrak{M},\text{ass}}(\exists\xi(P_0\xi \wedge \neg P_0\xi)) \in T_{\mathfrak{M}}$). In other words, in \mathbf{CT}^1 , the conjunction of the coordinates of the axiomatic base of \mathcal{N}^1 entails $\exists\xi(P_0\xi \wedge \neg P_0\xi)$ —the joint existence of positive and negative cases and the non-existence of a sharp boundary between them entails the existence of cases which are both positive and negative. Given (NEC), a theorist who would like to accept ψ ought not to accept that there are no cases which are both positive and negative. Indeed, given (NDD) (see section 5.4.3), a theorist who would like to accept ψ for non-soritical reasons (see again section 5.4.3) ought to accept that there are cases which are both positive and negative.

It is crucial to see that this concession is not as catastrophic as it might seem at first glance. For no δ do we have the entailment from the conjunction of the coordinates of the axiomatic base of \mathcal{N}^1 to $P_0\delta \wedge \neg P_0\delta$, since, even though, in every \mathbf{CT}^1 -model \mathfrak{M} of interest, for some δ , $P_0\delta \wedge \neg P_0\delta \in T_{\mathfrak{M}}$, it is not the case that, for some δ , in every \mathbf{CT}^1 -model \mathfrak{M} of interest, $P_0\delta \wedge \neg P_0\delta \in T_{\mathfrak{M}}$. What the conjunction of the coordinates of the axiomatic base of \mathcal{N}^1 logically requires is not the contradictoriness of any particular object, but the contradictoriness of some object or other. One can accept *that some object or other* is contradictory without accepting *of any particular object* that it is contradictory; in fact, one can accept that while *rejecting* of each particular object that it is contradictory and indeed *accepting* of each particular object that it is *not* contradictory—or, at least, given (PI), one can do so without

violating (AR) if one is accepting that some object or other is contradictory for soritical reasons (see section 5.4.3).

Say that a semantics is “*non-value-functional*” iff the value of some compound expressions is not a function of the values of the component expressions (or instances), and say that a semantics is “*non-truth-functional*” iff it fails to satisfy some entailments between the truth (falsity) of a compound expression and the truth or falsity of the component expressions (or instances). Our semantics are all value-functional and, whenever (SIMP_0) , (SIMP_1) , (ADD_0) , (ADD_1) , (ADJ) , (DISJ) , (NP) , (NC) , (UI) , (PG) , (UG) and (PI) are present, they are truth-functional as well. We can then recast the gist of the foregoing considerations on the existence of a contradictory object by saying that non-transitivity allows to simulate non-truth-functional effects in a truth-functional framework. Such a non-truth-functional effect can then be plausibly regarded as expressing the contradictoriness of a soritical series in its totality rather than the contradictoriness of any particular object, and the resistance against the former is arguably much weaker than the resistance against the latter.

Despite these suggestive remarks, theorem 4.5.11 certainly encourages the exploration of various strategies for weakening some of our semantic constraints. The heavy use of (DNE) made in the proof may for instance invite for an intuitionist rejection of $(\text{neg}_1^{\leftarrow})$. Even though they are undoubtedly worth investigating further, I think that this and other strategies focussing on the behaviour of negation and on its interaction with conjunction and disjunction fail to deal with the root of the problem we are confronting. For this root arguably lies in the crucial iterated use of distribution in the proof of theorem 4.5.11, parallelling that made in the proof of theorem 4.3.15. There, we managed to avoid paradox by rejecting (LTT^l) and (LTT^r) , but the foregoing result shows that such a rejection does not really go to the heart of the problem. In a nutshell, this consists in the fact that the strictness of $(\text{DISTR}_{\wedge/\vee})$ allows us to transform without loss of goodness the conjunction of a certain decision (in the positive or in the negative), for every object x in

some subseries S_0 of a soritical series, of the question whether x is P_0 with the decision (in the positive or in the negative), for another object y , of the question whether y is P_0 into a certain decision (in the positive or in the negative), for every object z in the subseries S_1 extending S_0 by y , of the question whether z is P_0 . The preservation of goodness allows this process to be iterated indefinitely, so that, in the end, a certain decision (in the positive or in the negative) is reached, for every object x in the soritical series in its totality, of the question whether x is P_0 , thereby contravening what is arguably an essential feature of soritical series—the peculiar *unsurveyability* which pertains to them in spite of their finitude (section 6.5.3 will develop this theme).

Distribution appears even more problematic once we consider that the only intuitive argument in its favour unavoidably appeals, among other things, to (DP). The argument is well-known from discussions of quantum logics and runs as follows (in the specific version justifying $(\text{DISTR}_{\wedge/\vee})$). Assume φ, ψ . By (ADJ), $\varphi \wedge \psi$, and so, by (ADD_0) and (T^l) , $(\varphi \wedge \psi) \vee (\varphi \wedge \chi)$ (the application of (T^l) is justified since (ADD_0) has the full 1-left-transitivity property). Assume φ, χ . By (ADJ), $\varphi \wedge \chi$, and so, by (ADD_1) and (T^l) , $(\varphi \wedge \psi) \vee (\varphi \wedge \chi)$ (the application of (T^l) is justified since (ADD_1) has the full 1-left-transitivity property). Therefore, by (DP), (C^l) , (W^l) and (W^r) , $\varphi, \psi \vee \chi$ entails $(\varphi \wedge \psi) \vee (\varphi \wedge \chi)$, and so, by (C^l) , (CP_0) and (CP_1) , $\varphi \wedge (\psi \vee \chi)$ entails $(\varphi \wedge \psi) \vee (\varphi \wedge \chi)$. The argument crucially uses (DP), which fails in every tolerant logic (note that theorem 4.3.13 does not apply in this case, since the sequences of the relevant argument forms do not have the full 1-right-transitivity property).

I thus conclude that we have good reasons to reject the semantic underpinning of the strictness of $(\text{DISTR}_{\wedge/\vee})$ and $(\text{DISTR}_{\vee/\wedge})$ —that is, $(\text{glb}/\text{lub}_1^2)$ (and $(\text{lub}/\text{glb}_1^2)$). Not only does this rejection free us of unwelcome logical strength, but it also comes at no additional cost, as the distributivity of our semantic structures is quite independent from their other features, so that all the other properties of tolerant logics we have been studying are preserved

even if distributivity is lost, as inspection of the various proofs readily reveals. Let us then consider the class of **NDCT**¹-structures (and, consequently, of **NDCT**¹-models) which satisfy all the constraints on **CT**¹-models save for possibly (glb/lub₁²) (and (lub/glb₁²)), **NDCT**¹ being the non-distributive “classical” tolerant logic generated by these structures. It must be shown that there is indeed a suitable **NDCT**¹-model of the axiomatic base of \mathcal{N}^1 and that such a model does not also tolerate any wff to the effect that some object or other is contradictory. For simplicity’s sake, we focus again on the consistency result for **NDCT**¹_→ (that is, the restriction of **NDCT**¹ to the extensional language $\mathcal{L}_{\rightarrow}^1$). The extension to full **NDCT**¹ is straightforward.

Theorem 4.5.12. *There is a **NDCT**¹_→-model \mathfrak{M} of the axiomatic base of \mathcal{N}^1 such that, for no wff φ of the form $\exists\xi_0\exists\xi_1\exists\xi_2\ldots\exists\xi_i(((\ldots(\psi_0 \wedge \neg\psi_0) \vee (\psi_1 \wedge \neg\psi_1) \vee (\psi_2 \wedge \neg\psi_2) \ldots (\psi_i \wedge \neg\psi_i)))$ (where each ψ_j may have the same form as φ) and assignment $\text{ass}, \text{val}_{\mathfrak{M}, \text{ass}}(\varphi) \in T_{\mathfrak{M}}$.*

Proof. We consider the **NDCT**¹_→-model \mathfrak{C}_1^1 where:

- $U_{\mathfrak{C}_1^1} = \{a_i : 0 \leq i \leq 1,000,000\} \cup \{b_i : 0 \leq i \leq 1,000,000\};$
- $V_{\mathfrak{C}_1^1} = \text{pow}(\{0, 1\}) \cup \{X : X \in \text{pow}(\{0, 1, 2, 3\}) \text{ and } \{0, 1\} \subseteq X\} \cup \{X : X \in \text{pow}(\{0, 1, 2, 3, 4, 5\}) \text{ and } \{0, 1, 2, 3\} \subseteq X\} \cup \{\{0, 4\}, \{1, 5\}\};$
- $D_{\mathfrak{C}_1^1} = \{X : X \in \text{pow}(\{0, 1, 2, 3, 4, 5\}) \text{ and } \{0, 1, 2, 3\} \subseteq X\};$
- $\preceq_{\mathfrak{C}_1^1} = \{\langle X, Y \rangle : X \subseteq Y\};$
- $\text{tol}_{\mathfrak{C}_1^1} = \{\langle X, Y \rangle : \text{either } Y = \{Z : \text{card}(Z) \geq \text{card}(X) - 1\} \text{ or } [X = \{0, 1, 2, 3\} \text{ and } Y = \{Z : \text{card}(Z) \geq 3\} \cup \{\{0, 4\}, \{1, 5\}\}]\};$
- $O_{\mathfrak{C}_1^1} = \{\text{neg}_{\mathfrak{C}_1^1}, \text{id}_{\mathfrak{C}_1^1}\}$, where $\text{neg}_{\mathfrak{C}_1^1} = \{\langle X, Y \rangle, \langle Y, X \rangle : \text{either } [\text{card}(X) = 6 \text{ and } \text{card}(Y) = 0] \text{ or } [X = \{0, 1, 2, 3, 4\} \text{ and } Y = \{1\}] \text{ or } [X = \{0, 1, 2, 3, 5\} \text{ and } Y = \{0\}] \text{ or } [X = \{0, 1, 2, 3\} \text{ and } Y = \{0, 1\}] \text{ or none of the foregoing conditions is met and } X \neq Y \text{ and } \text{card}(X) = \text{card}(Y)]$ and $\text{id}_{\mathfrak{C}_1^1}$ is such that, if $\text{id}_{\mathfrak{C}_1^1}(u_0, u_1) \in D_{\mathfrak{C}_1^1}$, then:

- (i) For every $\Phi^i \in PRED_{\mathcal{L}^1}^i$, if $[\text{card}(\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)) \geq 5 \text{ or } \leq 2]$ and $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) \neq \{0, 1\}$, then $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) [j, k : k - j = i]$;
- (ii) For every $\Phi^i \in PRED_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1, 2, 3\}$, then either $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{0, 1, 2\} [j, k : k - j = i]$;
- (iii) For every $\Phi^i \in PRED_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1, 2\}$, then either $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{0, 1\} [j, k : k - j = i]$;
- (iv) For every $\Phi^i \in PRED_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1, 3\}$, then either $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{0, 1, 2, 3\} [j, k : k - j = i]$;
- (v) For every $\Phi^i \in PRED_{\mathcal{L}^1}^i$, if $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k) = \{0, 1\}$, then either $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_0 \dots u_k)$ or $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\Phi^i)(u_j, u_{j+1}, u_{j+2} \dots u_1 \dots u_k) = \{0, 1, 3\} [j, k : k - j = i]$.

Moreover, $\text{id}_{\mathfrak{C}_1^1}$ is such that:

- (i') $\text{id}_{\mathfrak{C}_1^1}(u_0, u_1) = \{0, 1, 2, 3\}$ iff, either $u_0 = u_1$ or, for some i, j [$i : 0 \leq i, j \leq 1,000,000$], $u_0 = b_i$, $u_1 = b_j$ and $|i - j| \leq 1$;

- (ii') $\text{id}_{\mathfrak{C}_1^1}(u_0, u_1) = \{0, 1, 2\}$ iff, for some i, j [$i : 0 \leq i, j \leq 1,000,000$],
 $u_0 = b_i$, $u_1 = b_j$ and $|i - j| = 2$;
- (iii') Otherwise, $\text{id}_{\mathfrak{C}_1^1}(u_0, u_1) = \{0, 1\}$.

It's easy to check that these additional conditions on $\text{id}_{\mathfrak{C}_1^1}$ are consistent with conditions (i)-(v).

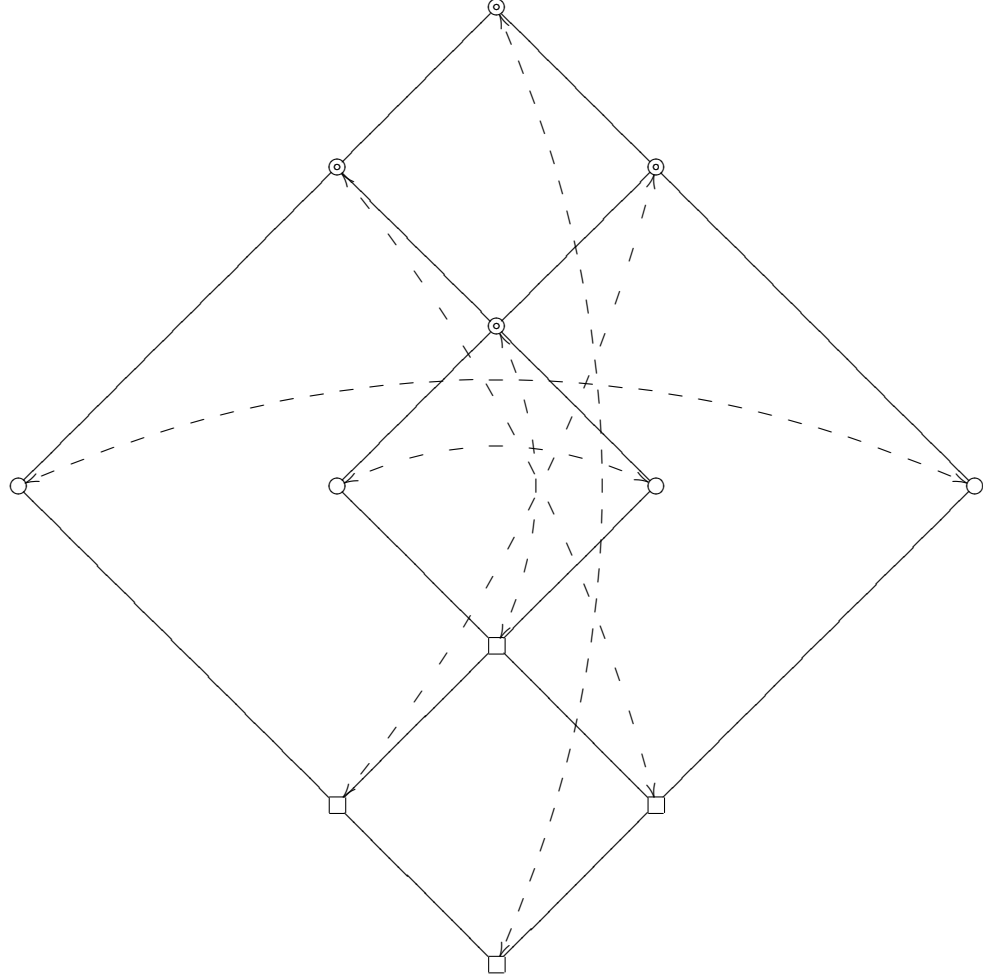
- $\text{int}_{\mathfrak{C}_1^1}$ is such that:

- (i) If, for some i [$i : 0 \leq i \leq 1,000,000$], $\delta = a_i$ or $\delta = b_i$, then
 $\text{int}_{\mathfrak{C}_1^1}(\delta) = \delta$;
- (ii) $\text{int}_{\mathfrak{C}_1^1}(@) = \{\langle \langle u_0 \rangle, u_1 \rangle : \text{for some } i [i : 0 \leq i \leq 1,000,000], \text{ either}$
 $[u_0 = a_i \text{ and } u_1 = b_i] \text{ or } [u_0 = b_i \text{ and } u_1 = a_i]\}$;
- (iii) $\text{int}_{\mathfrak{C}_1^1}(') = \{\langle \langle u_0 \rangle, u_1 \rangle : \text{for some } i [i : 0 \leq i < 1,000,000], \text{ either}$
 $[u_0 = a_i \text{ and } u_1 = a_{i+1}] \text{ or } [u_0 = a_{1,000,000} \text{ and } u_1 = a_{1,000,000}] \text{ or}$
 $[u_0 = b_i \text{ and } u_1 = b_{i+1}] \text{ or } [u_0 = b_{1,000,000} \text{ and } u_1 = b_{1,000,000}]\}$;
- (iv) $\text{int}_{\mathfrak{C}_1^1}(P_0) = \{\langle \langle u \rangle, v \rangle : \text{either, [for some } i [i : 0 \leq i \leq 499,998], u =$
 $a_i \text{ and } v = \{0, 1, 2, 3, 4, 5\}] \text{ or } [u = a_{499,999} \text{ and } v = \{0, 1, 2, 3, 4\}]$
 $\text{or } [u = a_{500,000} \text{ and } v = \{0, 4\}] \text{ or } [u = a_{500,001} \text{ and } v = \{0\}] \text{ or}$
 $[u = a_{500,002} \text{ and } v = \emptyset]; \text{ otherwise, } v = \emptyset\}$;
- (v) $\text{int}_{\mathfrak{C}_1^1}(R_0) = \{\langle \langle u_0, u_1 \rangle, v \rangle : \text{for some } i, j [i, j : 0 \leq i, j \leq$
 $1,000,000], u_0 = a_i \text{ and } u_1 = a_j \text{ and [either } [|i - j| \leq 1 \text{ and}$
 $v = \{0, 1, 2, 3\}] \text{ or } [|i - j| = 2 \text{ and } v = \{0, 1, 2\}] \text{ or } [|i - j| \geq 2 \text{ and}$
 $v = \{0, 1\}]\}$; otherwise, $v = \{1\}$.

It's easy to check that these additional conditions on P_0 and R_0 are consistent with conditions (i)-(v) on $\text{id}_{\mathfrak{C}_1^1}$.

It's easy to check that \mathfrak{C}_1^1 is indeed a $\mathbf{NDCT}_{\rightarrow}^1$ -model for the axiomatic base of \mathcal{N}^1 and is such that, for no wff φ of the form $\exists \xi_0 \exists \xi_1 \exists \xi_2 \dots \exists \xi_i (((\dots (\psi_0 \wedge \neg \psi_0) \vee (\psi_1 \wedge \neg \psi_1) \vee (\psi_2 \wedge \neg \psi_2) \dots (\psi_i \wedge \neg \psi_i)))$ (where each ψ_j may have the same form as φ) and assignment ass , $\text{val}_{\mathfrak{C}_1^1, \text{ass}}(\varphi) \in T_{\mathfrak{C}_1^1}$ (an easy induction on

the complexity of wffs building on the observation that, for every $v \in V_{\mathfrak{C}_1^1}$, $\text{glb}_{\mathfrak{C}_1^1}(\{v, \text{neg}_{\mathfrak{C}_1^1}(v)\}) \preceq_{\mathfrak{C}_1^1} \{0, 1\} \in V_{\mathfrak{C}_1^1} \setminus T_{\mathfrak{C}_1^1}$. \mathfrak{C}_1^1 may be depicted by the following Hasse diagram (notational conventions as in proof 4.3.10):



□

Analogously to \mathfrak{C}_0^1 , not only does a model such as \mathfrak{C}_1^1 provide a proof of the consistency of the axiomatic base of \mathcal{N}^1 in $\mathbf{NDCT}_{\rightarrow}^1$, but it also validates the other traditional fundamental rule governing identity, (II).

Definition 4.5.14. A \mathbf{NDCT}^1 -model \mathfrak{M} is *quasi- \mathfrak{C}_1^1 -equivalent* iff there exists an isomorphism between the structure of \mathfrak{M} and the structure of \mathfrak{C}_1^1 possibly with the exception of conditions (i')-(iii') on $\text{id}_{\mathfrak{C}_1^1}$.

Theorem 4.5.13. (II) is valid in every quasi- \mathfrak{C}_1^1 -equivalent \mathbf{NDCT}^1 -model.

Proof. An easy, even though tedious, induction on the complexity of wffs for conditions (i)-(v) on id.

□

Again, note that (II) is licensed by the conditions (i)-(v) on id, which in effect amount to requiring, informally (and plausibly), that, if both the identification of x with y and the information that y is F are very good, then the information that x is F is at least good enough. Given (II)'s eminent plausibility, I propose that we single out for special attention the strengthening \mathbf{NDCIIT}^1 of \mathbf{NDCT}^1 characterized by quasi- \mathfrak{C}_1^1 -equivalent \mathbf{NDCT}^1 -models.

As with \mathfrak{C}_0^1 , other features of \mathfrak{C}_1^1 are also worth remarking upon. Not only does \mathfrak{C}_1^1 validate $(N^{1a\Rightarrow})$ and $(N^{1a\Leftarrow})$, it also validates the strictness of the argument from $@\tau_0 \simeq @\tau_1$ to $R_0\tau_0\tau_1$ and *vice versa*. This can easily be seen as requiring that [the relation denoted by R_0 is transitive iff the relation denoted by \simeq is transitive], and in \mathfrak{C}_1^1 they are in effect both transitive (even though, of course, not chain-transitive).

Furthermore, \mathfrak{C}_1^1 is such that, unlike many other \mathbf{NDCT}^1 -models, for every $X \subseteq D_{\mathfrak{C}_1^1}$, $\text{glb}(X) \in D_{\mathfrak{C}_1^1}$. This implies that every quasi- \mathfrak{C}_1^1 -equivalent \mathbf{NDCT}^1 -model of $\Gamma = \langle \varphi_0, \varphi_1, \varphi_2 \dots \varphi_i \rangle$ is also a model of $\varphi_0 \wedge \varphi_1 \wedge \varphi_2 \dots \wedge \varphi_i$, so that the consistency in \mathbf{NDCIIT}^1 of the (finite) axiomatic base of a theory \mathcal{T} implies the consistency in \mathbf{NDCIIT}^1 of the conjunction of the axioms of \mathcal{T} (and the consistency in \mathbf{NDCIIT}^1 of the axiomatic base $\langle \varphi[\tau_0/\xi], \varphi[\tau_1/\xi], \varphi[\tau_2/\xi] \dots \rangle$ (where ξ does not occur bound in φ) of a theory \mathcal{T} implies the consistency in \mathbf{NDCIIT}^1 of $\forall \xi \varphi$). Dually, \mathfrak{C}_1^1 is such that, unlike many other \mathbf{NDCT}^1 -models, for every $X \subseteq V_{\mathfrak{C}_1^1} \setminus T_{\mathfrak{C}_1^1}$, $\text{lub}(X) \in V_{\mathfrak{C}_1^1} \setminus T_{\mathfrak{C}_1^1}$. This implies that every quasi- \mathfrak{C}_1^1 -equivalent \mathbf{NDCT}^1 -model which does not tolerate $\Gamma = \langle \varphi_0, \varphi_1, \varphi_2 \dots \varphi_i \rangle$ does not tolerate $\varphi_0 \vee \varphi_1 \vee \varphi_2 \dots \vee \varphi_i$ either, so that the non-tolerance in \mathbf{NDCIIT}^1 of the (finite) axiomatic base of a theory \mathcal{T} implies the non-tolerance in \mathbf{NDCIIT}^1 of the disjunction of the

axioms of \mathcal{T} (and the non-toleration in \mathbf{NDCIIT}^1 of the axiomatic base $\langle \varphi[\tau_0/\xi], \varphi[\tau_1/\xi], \varphi[\tau_2/\xi] \dots \rangle$ (where ξ does not occur bound in φ) of a theory \mathcal{T} implies the non-toleration in \mathbf{NDCIIT}^1 of $\exists \xi \varphi$).

Finally, despite the failure of distributivity in \mathbf{NDCIIT}^1 , we still have the peculiar incompleteness phenomenon remarked upon in section 4.5.7:

Theorem 4.5.14. *\mathbf{NDCIIT}^1 does not have the Lindenbaum property.*

Proof. Consider $\Gamma = \langle (N^{1^p}), (N^{1^n}), (N^{1^t}), (N^{1^{a^n}}), (N^{1^{a^t}}), (N^{1^{a^\Rightarrow}}), (N^{1^{a^\Leftarrow}}) \rangle$. As proved in theorem 4.5.12, $\Gamma \not\vdash_{\mathbf{NDCIIT}^1} \emptyset$. However, for no Δ such that, for every φ , if $\varphi \in \text{ran}(\Gamma)$, then $\varphi \in \text{ran}(\Delta)$, and, for every i [$i : 0 \leq i \leq 1,000,000$], either Pa_i or $\neg Pa_i \in \text{ran}(\Delta)$, $\Delta \not\vdash_{\mathbf{NDCIIT}^1} \emptyset$. For, since $(N^{1^p}), (N^{1^n}) \in \text{ran}(\Gamma)$, it would follow by familiar reasoning (in the classical metalanguage) that, for some i , $Pa_i \in \text{ran}(\Delta)$ and $\neg Pa_{i+1} \in \text{ran}(\Delta)$. But, since $(N^{1^t}) \in \text{ran}(\Gamma)$, this would be impossible, as $\varphi, \psi, \neg \exists \xi (\varphi \wedge \psi) \vdash_{\mathbf{NDCIIT}^1} \emptyset$ ((ADJ) has the full 0-left-transitivity property in \mathbf{NDCIIT}^1).

□

4.6 Conclusion

We have thus seen how a fairly natural weakening of the logic, targeting one of the structural properties of the consequence relation rather than any property pertaining to a specific logical constant, is sufficient to stabilize the first-order fragment of the naive theory of vagueness. I think that this result is crucial in reinstating the naive theory as one of the main competitors in the vagueness debate. Distinctively technical issues—such as the extension of tolerant logics to higher orders, the development of adequate deductive systems for the logics and the possibility of an alternative possible-world semantics for them—will have to be investigated elsewhere. In the next chapter, we rather turn to some distinctively philosophical issues arising from the non-transitivist solution to the sorites paradox presented here.

Chapter 5

Following-from and Transitivity

5.1 Introduction and Overview

I start with some rather dogmatic statements, simply in order to fix a specific enough framework against which to investigate the topic of this chapter. The reader who does not share some or all of the doctrines thereby expressed is invited to modify the rest of the discussion in this chapter in accordance to her favourite views on logical consequence.

Sometimes, some things logically *follow from* some things. The former are a logical *consequence* of the latter and, conversely, the latter logically *entail* the former. Something logically following “from nothing” is a *logical truth*.¹ There are presumably other ways of following-from (conceptually, metaphysically, nomologically etc.). Even within the logical way of following-from, this can be determined by different features of the sentences in question. Assuming a *semantic* individuation of sentences (two sentences are the same only if they mean exactly the same things), the relevant features can be:

- The *semantic value* of some expressions or others (together with syntactic structure and the identities of all occurring expressions), as in

¹For readability’s sake, I will henceforth mostly drop the qualification ‘logical’.

‘John is unmarried’'s following from ‘John is a bachelor’;

- The semantic value of some expressions belonging to a *privileged class* of “*logical constants*” (again together with syntactic structure and the identities of all occurring expressions), as in ‘Snow is white’'s following from ‘Snow is white and grass is green’;
- The *semantic structure* of the sentences, determined by the semantic categories of the expressions occurring in them and their mode of composition (together with the identities of all occurring expressions), as in ‘New York is a city’'s following from ‘New York is a great city’ (see Evans [1978] and Sainsbury [2001], pp. 359–64 for some discussion);
- The sheer *identities* of the sentences, as in ‘Snow is white’'s following from ‘Snow is white’ (see Varzi [2002], pp. 213–4; Moruzzi and Zardini [2007], pp. 180–2 for critical discussion).

The last three cases are usually considered to be cases of “*formal consequence*”. In the following, we focus attention on them, even though, as should have been clear already from some remarks in section 2.1, on my view the interesting divide between logical following-from and the rest does not coincide with the divide between formal logical following-from and the rest, the latter simply arising from a division of the theoretically basic notion of logical following-from into subnotions which are individuated by the specific features of the sentences which, in each particular case, make it so that the relation of following-from obtains (see García-Carpintero [1993] for a similar view). Unfortunately, I won’t have anything more to say here about either the basic notion of logical following-from or the derivative notion of formal logical following-from (see Moruzzi and Zardini [2007], pp. 161–74 for a critical survey of the main approaches to the analysis of logical following-from and formal logical following-from).

The things related by consequence are *sentences*. No interesting notion of a proposition as a genuinely non-linguistic entity has been made out which

would sustain the fine-grainedness of individuation required by consequence. ‘Barbarelli is Giorgione’ expresses the same proposition as ‘Barbarelli is Barbarelli’ on many accounts of propositions, but only the latter is a logical truth. ‘The US is the United States’ expresses the same proposition as ‘The US is the US’ on any account of propositions worthy of consideration, but only the latter is a (formal) logical truth. Utterances are also ill-suited to play the relevant role in consequence, as utterances of presumed logical truths such as ‘Here is here’ can be uttered falsely (if two different places are indicated in association with the two different occurrences of ‘here’). Moreover, sentences are nowadays the objects invariably considered by the science of consequence, logic (see Moruzzi and Zardini [2007], pp. 179–80 for further discussion).

In consequence, sentences are often “put together” (what this amounts to in the case of tolerant logics has been seen in section 4.4.1). Their mode of being “put together” is signalled in English by ‘and’ in the locution ‘ $‘Q_0’$ and $‘Q_1’$ and $‘Q_2’ \dots$ follow from $‘P_0’$ and $‘P_1’$ and $‘P_2’ \dots$ ’ and cannot be assumed to be less structured than the mode of being put together enjoyed by the coordinates of a sequence. This is so because many logics give divergent answers to the questions:

- (i) whether ψ follows from φ ;
- (ii) whether ψ follows from φ and φ .

But φ and φ form the same plurality, set, compound, aggregate, fusion etc. as φ . Some logics even discriminate between:

- (i’) χ ’s following from φ and ψ ;
- (ii’) χ ’s following from ψ and φ .

But φ and ψ form the same multi-set as ψ and φ (see Restall [2000] and Paoli [2002] for useful overviews of such logics). We will thus stick to the

practice of chapter 4 of representing with sequences such a fine-grained mode of “putting together” sentences.

In such a framework, consequence plays such a central role that all other traditional logical properties can be defined in terms of it. To give some examples, Γ is *consistent* iff \emptyset does not follow from it; Γ is *exhaustive* iff it follows from \emptyset ; Γ and Δ are *contrary* iff \emptyset follows from Γ, Δ ; Γ and Δ are *subcontrary* iff Γ, Δ follows from \emptyset . Γ and Δ are *compatible* iff they are not contrary; Γ and Δ are *contradictory* iff they are both contrary and subcontrary.

An *argument* is a structure representing some or no sentences (the *conclusions* of the argument) as following from some or no sentences (the *premises* of the argument). In English, an argument is usually expressed with a discourse of the form ‘ P_0 ; (*and*) P_1 ; (*and*) $P_2 \dots$ Therefore, Q_0 ; (*or*) Q_1 ; (*or*) $Q_2 \dots$ ’. An argument is *valid* iff the conclusions in effect follow from the premises; *sound* iff it is valid and all of its premises are true. An *inference* is an act of drawing conclusions from premises (I should stress for further reference that I am not equating inferring with drawing conclusions in accordance to a certain collection of syntactic rules). A *deduction* is an abstract codification of a derivation of conclusions from premises conforming to a certain collection of syntactic rules (deductive rules). A *proof* is a deduction from 0 premises.

What does it mean to draw (accept, reject, doubt etc.) conclusions when they are not exactly one conclusion, but many or none (as it might be the case in our general multiple-conclusion framework)? As a first approximation, we can say that, while premises have to be treated (accepted, rejected, doubted etc.) “*conjunctively*”, conclusions have to be treated (accepted, rejected, doubted etc.) “*disjunctively*”. Focussing on acceptance, it is important to note that accepting “disjunctively” a sequence should not be interpreted as accepting every (or even some) coordinate of the sequence. For example, in most multiple-conclusion logics, ⟨‘There are 1,963 houses in St Andrews’, ‘It is not the case that there are 1,963 houses in St Andrews’⟩ follows from

‘Either there are 1,963 houses in St Andrews or it is not the case that there are 1,963 houses in St Andrews’, and in most of these logics one may well rationally accept the latter while having no idea of how many houses there are in St Andrews. If one is in such a situation, it would of course be wrong not only to conclude that one is committed to accepting *both* ‘There are 1,963 houses in St Andrews’ *and* ‘It is not the case that there are 1,963 houses in St Andrews’, but also to conclude that *either* one is committed to accepting ‘There are 1,963 houses in St Andrews’ *or* one is committed to accepting ‘It is not the case that there are 1,963 houses in St Andrews’. One is only committed as it were “disjunctively” to accepting ‘There are 1,963 houses in St Andrews’ and ‘It is not the case that there are 1,963 houses in St Andrews’, which can very roughly be characterized as a commitment to accepting that either ‘There are 1,963 houses in St Andrews’ is true or ‘It is not the case that there are 1,963 houses in St Andrews’ is true.²

A major part of the philosophical investigation of the notion of consequence consists in an attempt at elucidating its *nature*—what consequence consists in. Yet, consequence is also a relation, and as such one can sensibly ask what its *formal properties* are.³ Arguably, Tarski’s most notorious contribution to the philosophical investigation of the notion of consequence is constituted by his theory of what consequence consists in: truth preservation in every model (see Tarski [1936]). An at least equally important contribution to such investigation is however represented by his earlier studies concerning an abstract theory of consequence relations, aimed at determining the for-

²In this connection, I should also note that accepting (rejecting, doubting etc.) a sequence can only be equated with accepting (rejecting, doubting etc.) all (or some) of the sentences in its range if the underlying logic exhibits (C^l) , (C^r) , (W^l) and (W^r) (as all tolerant logics do).

³Compare with resemblance: a major task for resemblance theories is to determine what resemblance between two individuals consists in (sharing of universals, matching in tropes, primitive similarity etc.); yet, the study of the formal properties of resemblance (seriality, reflexivity, symmetry etc.) can fruitfully be pursued even in the absence of an answer to the question about its ultimate nature.

mal behaviour of any such relation. In his Tarski [1930], he mentions four properties a consequence relation worthy of this name must have: reflexivity, monotonicity, transitivity and (less centrally) compactness.⁴ I think all of these properties are at least questionable (see Moruzzi and Zardini [2007], pp. 180–7). But here I want to focus on transitivity, trying to make sense of a position according to which consequence is not transitive (see chapter 4 for a precise statement of the main transitivity properties). I will not argue for any such position (I did so in chapter 4), but will simply try to make it adequately intelligible and assess what impact its correctness would have on our understanding of consequence.

The rest of the chapter is organized as follows. To fix ideas, section 2 puts on the table a range of philosophically interesting non-transitive consequence relations, introducing briefly their rationale. Section 3 discusses and disposes of two very influential objections of principle to the use of non-transitive consequence relations. Section 4 delves into some fine details of the logical and normative structures generated by non-transitivity. Section 5 draws the conclusions which follow from these investigations for our general understanding of the relation between logical consequence and rationality and for the non-transitivist solution to the sorites paradox.

5.2 Non-Transitive Consequence Relations

5.2.1 Relevance

In order to study in details the impacts of non-transitivity on our understanding of consequence, it will be helpful to have in mind concrete examples of consequence relations lacking the transitivity property—the specific reasons

⁴Important as they may be, it is worth stressing that they would still grossly underdetermine the identification of the nature of consequence, even if the field of the relation is kept fixed. Consider e.g. that the relation which holds between Γ and Δ iff Γ entails Δ and is non-empty will satisfy all the properties mentioned in the text if consequence does.

underlying their rejection of transitivity will help us to form and test an understanding of general features of consequence which opens daylight for the idea of its failing to be transitive.⁵ Of course, again, the aim here is not to argue for the adoption of any such logic, but only to shed light on their rational motivation. Let me also anticipate that the rest of our discussion will actually almost exclusively centre on left transitivity, as this is the target of all extant non-transitive logics.

Consider the two following arguments:

CONTRADICTION Intuitively, ‘Graham Priest is dead wrong’ does not follow from ‘The Strengthened-Liar sentence is true and the Strengthened-Liar sentence is not true’. Graham Priest’s error should not be entailed by the correctness of one of his most famous doctrines. Yet:

- (i) Both ‘The Strengthened-Liar sentence is true’ and ‘The Strengthened-Liar sentence is not true’ do seem to follow from it (by simplification);
- (ii) ‘The Strengthened-Liar sentence is true or Graham Priest is wrong’ does seem to follow from ‘The Strengthened-Liar sentence is true’ (by addition);
- (iii) ‘Graham Priest is dead wrong’ does seem to follow from ‘The Strengthened-Liar sentence is true or Graham Priest is wrong’ and ‘The Strengthened-Liar sentence is not true’ (by disjunctive syllogism).

Two applications of (T^l) (see section 4.2.3) would then yield that ‘Graham Priest is dead wrong’ does after all follow from ‘The Strengthened-Liar sentence is true and the Strengthened-Liar sentence is not true’.

⁵The reader should keep in mind that, throughout, what is meant by ‘failure of transitivity’ and the like is simply failure of *unrestricted* transitivity and the like. The most interesting non-transitive logics retain transitivity for many argument forms (see section 4.2.3 for a precise formulation of what this amounts to).

LOGICAL TRUTH Intuitively, ‘Timothy is a male and Timothy is a sibling iff it is not the case that [Timothy is not a male or Timothy is not a sibling]’ does not follow from ‘Timothy is a brother iff [Timothy is a male and Timothy is a sibling]’. No De Morgan Law should be entailed by an analysis of ‘brother’. Yet:

- (i’) ‘Timothy is a brother iff it is not the case that [Timothy is not a male or Timothy is not a sibling]’ does seem to follow from ‘Timothy is a brother iff [Timothy is a male and Timothy is a sibling]’ and ‘Timothy is a male and Timothy is a sibling iff it is not the case that [Timothy is not a male or Timothy is not a sibling]’ (by transitivity of the conditional), and so from the former only (by suppression of logical truths);
- (ii’) ‘Timothy is a brother iff [Timothy is a male and Timothy is a sibling]’ does seem to follow from itself (by reflexivity). Given this and (i’), it also does seem that ‘Timothy is a brother iff [Timothy is a male and Timothy is a sibling]’ entails both ‘Timothy is a brother iff [Timothy is a male and Timothy is a sibling]’ and ‘Timothy is a brother iff it is not the case that [Timothy is not a male or Timothy is not a sibling]’ (by contraction);
- (iii’) Moreover, it does seem that the latter two jointly entail ‘Timothy is a male and Timothy is a sibling iff it is not the case that [Timothy is not a male or Timothy is not a sibling]’ (by transitivity of the conditional).

One application of (T^l) would then yield that ‘Timothy is a male and Timothy is a sibling iff it is not the case that [Timothy is not a male or Timothy is not a sibling]’ does after all follow from ‘Timothy is a brother iff [Timothy is a male and Timothy is a sibling]’.

Some (Bolzano [1837] (according to George [1983]; George [1986]) and then Lewy [1958], pp. 123–32; Geach [1958]; Smiley [1959], pp. 238–43; Wal-

ton [1979]; Epstein [1979]; Tennant [1987], pp. 185–200, 253–65)⁶ have taken these intuitive judgements at face value and concluded that (T^l) does not unrestrictedly hold (see Lewy [1976], pp. 126–31; Routley et al. [1982], pp. 74–8 for critical discussion of this approach). One possible way of elaborating the rationale for these judgements would be as follows (see von Wright [1957], pp. 175, 177). (Non-0-premise, non-0-conclusion) consequence is a relation holding in virtue of some *meaning connection* between premises and conclusions—its holding can never be determined merely by the independent logical status of some premises or conclusions. Thus, (non-0-premise, non-0-conclusion) consequence should never discriminate between logical truths and logical contingencies.

This intuitive constraint can then be made precise as the requirement that an argument is valid iff it can be obtained by substitution of sentences for atomic sentences from a valid argument none of whose premises or conclusions are logical truths or falsehoods (see Smiley [1959], p. 240). The constraint yields the desired results: it can easily be checked that, subject to a certain qualification concerning **LOGICAL TRUTH**, each subargument of the previous arguments satisfies the requirement, even though the overall arguments do not.

In particular, as for **CONTRADICTION**, notice that the two subarguments in (i) are valid, as they can be obtained from the valid argument ‘Snow is white and grass is green. Therefore, snow is white (grass is green)’ by substitution of ‘The Strengthened-Liar sentence is true’ for ‘Snow is white’ and of ‘The Strengthened-Liar sentence is not true’ for ‘Grass is green’; the other subarguments satisfy the requirement already in their present form. As for **LOGICAL TRUTH**, notice that the subargument in (iii’) is valid, as it can be obtained from the valid argument ‘Snow is white iff grass is green; snow is white iff water is blue. Therefore, grass is green iff water is blue’

⁶Sylvan [2000], pp. 47–9, 98–9 intriguingly mentions some possible medieval and early-modern sources, but, to the best of my knowledge, a satisfactory investigation into the pre-modern history of this logical tradition has yet to be undertaken.

by substitution of ‘Timothy is a brother’ for ‘Snow is white’, of ‘Timothy is a male and Timothy is a sibling’ for ‘Grass is green’ and of ‘It is not the case that [Timothy is not a male or Timothy is not a sibling]’ for ‘Water is blue’; the other subarguments considered as a whole single subargument satisfy the requirement already in their present form. Given what counts as meaning connection in this framework, addition of (T^l) would lead to a gross overgeneration of meaning connections between sentences—the genuine intensional dependencies between the premises and conclusions of each subargument would overgenerate into the bogus intensional dependencies between the premises and conclusions of the overall arguments.

It is often claimed that the imposition of this and similar additional constraints on the consequence relation amounts to changing the subject matter of logic, and does not really engage with the (self-proclaimed) traditional view according to which “*truth preservation*” is what consequence is all about. Note that such a claim cannot be addressed in the particular case of **CONTRADICTION** by holding that some contradictions (sentences of the form ‘ P and it is not the case that P ’) may be true whilst not everything is true, so that the argument would fail to be truth preserving in the straightforward sense of having true premises and false conclusions—for the truth of only some contradictions would presumably invalidate subargument (iii) (*qua* not truth preserving in the straightforward sense) and therefore prevent a possible failure of transitivity (see Priest [2006a], pp. 110–22).

The claim is however highly dubious on other grounds. For it is plausible to assume that there is a notion of conditionality (expressible in the language by \rightarrow) such that, if Δ, ψ follows from Γ, φ , then $\Delta, \varphi \rightarrow \psi$ follows from Γ . At least on some readings, ‘If φ , then ψ ’ does presumably express such a notion in English. But, on such a reading, ‘If the Strengthened-Liar sentence is true and the Strengthened-Liar sentence is not true, then Graham Priest is dead wrong’ seems to be false (indeed false-only) if anything is. Under the present assumption, it would, however, be a logical truth if **CONTRADICTION** were valid. Even though, we may assume, truth preserving in

the straightforward sense of not having true premises and false conclusions, such an argument would thus not be truth preserving in the only slightly less straightforward sense of being such that its validity would imply the validity of arguments which are not truth preserving in the straightforward sense. Under the present assumption about the behaviour of the conditional, there is thus a perfectly good sense in which, for someone attracted by the approach just sketched, licensing the validity of **CONTRADICTION**, transitivity would lead indeed to failures of truth preservation—what consequence is supposed to be all about (see section 5.3.2 for more on transitivity and truth preservation and Read [1981]; Read [2003] for a different reply on the issue of relevance and truth preservation).

5.2.2 Tolerance

This is of course familiar ground by now, but let us rehearse the gist of the non-transitivist solution to the sorites paradox from a perspective which will facilitate further philosophical reflection on it. Consider the premises:

- (1) A man with 0 hairs is bald;
- (2) A man with 1,000,000 hairs is not bald;
- (3) If a man with i hairs is bald, so is a man with $i + 1$ hairs.

All these premises are intuitively true, and, presumably *a fortiori*, consistent. However, from (3) we have that, if a man with 0 hairs is bald, so is a man with 1 hair, which, together with (1), yields that a man with 1 hair is bald. Yet, from (3) we also have that, if a man with 1 hair is bald, so is a man with 2 hairs, which, together with the previous lemma that a man with 1 hair is bald, yields that a man with 2 hairs is bald. With another 999,997 structurally identical arguments, we reach the conclusion that a man with 999,999 hairs is bald. From (3) we also have that, if a man with 999,999 hairs is bald, so is a man with 1,000,000 hairs, which, together with the

previous lemma that a man with 999,999 hairs is bald, yields that a man with 1,000,000 hairs is bald. 999,999 applications of (T^l) would then yield that the contradictory of (2) follows simply from (1) and (3).

In chapter 4 we have taken these intuitive judgements at face value and concluded that (T^l) does not unrestrictedly hold (Weir [1998], pp. 792–4; Béziau [2006] also briefly entertain this possibility). One possible way of elaborating the rationale for these judgements would be as follows. A major point of a vague predicate is to draw a difference in application between some cases which are far apart enough on a dimension of comparison relevant for the application of the predicate. The predicate should *discriminate* between some such cases. Hence, (1) and (2) must be enforced. Still, another major point of a vague predicate is not to draw any difference in application (from a true application to anything falling short of that) between any two cases which are close enough on a dimension of comparison relevant for the application of the predicate. The predicate should *not discriminate* between any two such cases. Hence, (3) must be enforced. Moreover, instances of (3) should allow *modus ponens*: what substance is there to the idea that there is no sharp boundary between i and $i + 1$ in matters of baldness if, given the premise that a man with i hairs is bald, I cannot detach and infer that a man with $i + 1$ hairs is bald? Given what counts as indiscriminability connection in this framework and given the fact that the topology and metric of the dimension allow for chains of close enough items whose extreme are far apart, addition of (T^l) would lead to a gross overgeneration of indiscriminability connections between sentences—the correct mandate of not drawing any unit-sized difference would overgenerate into the incorrect mandate of not drawing any 1,000,000-sized difference.

5.2.3 Probabilistic Reasoning

Finally, I would also like to put on the table a case of a non-deductive, *defeasible* consequence relation—that is, roughly, a relation which is supposed

to hold between premises and conclusions iff the truth of some of the latter is *reasonable* in the lights of the (partial) state of information represented by the former, even if not *guaranteed* by their truth (as is supposed to be the case for a deductive consequence relation). For example, the inference of ‘Al is a native speaker of Italian’ from ‘Al was born in Little Italy’ is eminently reasonable, as well as the inference of ‘Al was born in Italy’ from ‘Al is a native speaker of Italian’. However, the inference of ‘Al was born in Italy’ from ‘Al was born in Little Italy’ is eminently unreasonable. Under this intuitive understanding of defeasible consequence:

- (1′) ‘Al is a native speaker of Italian’ is a consequence of ‘Al was born in Little Italy’;
- (2′) ‘Al was born in Italy’ is a consequence of ‘Al is a native speaker of Italian’;
- (3′) ‘Al was born in Italy’ is not a consequence of ‘Al was born in Little Italy’,

whereas, given (1′) and (2′), (T^l) would rule out (3′).

One possible way of elaborating the rationale for these judgements would be as follows. Whereas on all *probability distributions* reasonable in the light of how things actually are the conditional probabilities of ‘Al is a native speaker of Italian’ and ‘Al was born in Italy’ on ‘Al was born in Little Italy’ and ‘Al is a native speaker of Italian’ respectively are both very high, the conditional probability of ‘Al was born in Italy’ on ‘Al was born in Little Italy’ is very low (if not = 0!).

The idea can be made more precise in different specific ways. Here is a fairly general recipe. Let a model \mathfrak{M} of a language \mathcal{L} be a probability distribution on the sentences of \mathcal{L} . Let the conditional probability functions be totally defined—assume to that effect a suitable probability calculus (for well-known examples, see Popper [1959]; Rényi [1970]). The probability distribution corresponding to \mathfrak{M} will thus be in effect determined by the set of

conditional probability functions—the unconditional probability in \mathfrak{M} of φ being the conditional probability in \mathfrak{M} of φ on ψ , for some logical truth ψ (assuming a suitable logic in the characterization of the probability calculus). Define then Δ to be a δ -consequence of Γ in \mathfrak{M} iff the conditional probability in \mathfrak{M} of the disjunction of all the coordinates of Δ on the conjunction of all the coordinates of Γ is $\geq \delta$. Finally, define Δ to be a δ -consequence of Γ iff, for every model \mathfrak{M} , Δ is a δ -consequence of Γ in \mathfrak{M} . Assuming that the logic featuring in the characterization of the probability calculus is classical, it's easy to check that, as defined, for every $\delta > 0$, δ -consequence is just classical consequence (for $\delta = 0$, δ -consequence is trivial). The desired supra-classical strength comes of course by restricting the range of admissible models to a set of (contextually determined) “reasonable” probability distributions.

No doubt this recipe still leaves a lot of leeway in the choice of the probability calculus and of the appropriate restrictions on models, and some of its parameters (such as the logic featuring in the characterization of the calculus) could interestingly be modified. However, it seems plausible that, whichever specific implementation is eventually chosen, the consequence relations so obtained will have a decent claim to codify at least in part the (contextually determined) canon of non-deductive, defeasible reasoning. If so, such a canon will not satisfy (T^I) : given what counts as evidential connection in this framework, addition of (T^I) would lead to a gross overgeneration of evidential connections between sentences—the reasonable rules of thumb that people born in Little Italy (typically) speak Italian and that Italian speakers were (typically) born in Italy would overgenerate into the unreasonable rule of thumb that people born in Little Italy were (typically) born in Italy.

It might legitimately be wondered what the point is of introducing a defeasible consequence relation in the course of an attempt at understanding the idea that deductive consequence is non-transitive. Yet, it will turn out that one of the gateways to this understanding is constituted by an appreciation of the normative import of consequence on rational attitudes. The non-transitivist can be made sense of as interpreting such import in a par-

ticular way. However, that import is not of course a privilege of deductive consequence, and so study of defeasible consequence relations with certain properties (such as the failure of transitivity) may well shed light on some features of deductive consequence relations with the same properties.⁷

5.3 Two Objections to the Very Idea of Non-Transitive Consequence

5.3.1 Consequence and Inference

I can see (and have encountered, in print and conversation) two main objections concerning the very idea of non-transitive consequence, objections which, if correct, would seem to doom from the start any interesting use of a non-transitive consequence relation. Their rebuttal will help to dispel some misunderstandings of what non-transitivity of consequence amounts to. A more positive characterization will be offered in the next section.

There should be an uncontroversial sense in which logic is oftentimes substantially *informative*. There should be an uncontroversial sense in which reasoning oftentimes leads to the *discovery* of new truths. For example, there should be an uncontroversial sense in which the derivation from a standard arithmetical axiom system that there are infinitely many prime numbers is rightly regarded as a substantial discovery about natural numbers, accomplished by purely logical means (under the assumption of the truth of the axioms), no matter what one's views are about the ultimate aptness of logi-

⁷Indeed, there seem to be some deep connections between probabilistic reasoning and at least some kinds of non-transitive deductive consequence relations. For example, the main logics of chapter 4 fail to satisfy some principles (such as (IP_▷), (DP), the 1-left-transitivity property for (ADJ) etc.) which are also invalid from a probabilistic point of view, even though they are valid on many non-probabilistic codifications of defeasible consequence, such as mainstream non-monotonic logics (see Makinson [2005] for a useful overview of these).

cal vocabulary to represent features of the world. However, as Quine among others has stressed (see e.g. Quine [1986], pp. 80–94), the elementary steps of logic are, in a sense, *obvious*.⁸ In what sense can then logic still be substantially informative?

Well, as the traditional thought about this puzzle has always been at pains to stress, a series of completely obvious elementary steps may well lead to a completely unobvious conclusion. The relation *x-is-an-obvious-consequence-of-y* is non-transitive. However, it is important to note that this by itself does not yet explain away the puzzle: a series of one-foot steps may well lead to cover a considerable distance, but no one has ever supposed this to show that one can take a much longer step than the length of one's legs would give reason to suppose. *x-is-obviously-reachable-in-one-step-from-y* is indeed non-transitive, but so is *x-is-reachable-in-one-step-from-y*. The crucial, implicit, auxiliary assumption must be that consequence, as opposed to obvious consequence, is indeed transitive, so that the path to the unobvious conclusion can be obliterated and this be seen already to follow from the initial assumptions. As Timothy Smiley so nicely put it, “the whole point of logic as an instrument, and the way in which it brings us new knowledge, lies in the contrast between the transitivity of ‘entails’ and the non-transitivity of ‘obviously entails’” (Smiley [1959], p. 242). Thus, there seems to be little space for consequence not to be transitive, if logic is to preserve its function as a means of discovering new truths.

I accept that logic is to preserve such a function, but, as it stands, I contest the very coherence of the previous argument. Suppose that ψ obviously follows from φ and that χ obviously follows from ψ but not so obviously from

⁸Quine uses ‘obvious’ in such a way that an apparently deviant logician would be best translated non-homophonically (see also Quine [1960], pp. 57–61). I find very dubious that even the elementary steps of logic are “obvious” in this sense and don’t mean anything that strong—only that once one has adopted a reasonable logic, its elementary steps look most straightforward and least informative, unlike the step from a standard arithmetical axiom system to Euclid’s theorem. Thanks to Graham Priest and Stewart Shapiro for stimulating discussions of Quine’s views on these matters.

φ . If consequence is indeed transitive, then, at least as far as consequence is concerned, the step from φ to χ is no more mediated than those from φ to ψ and from ψ to χ : χ just follows from φ as it follows from ψ (and as ψ follows from φ). If the step from φ to χ is indeed unobvious, then it is just not true, in a transitivist framework, that every “elementary” logical step is obvious. Thus, the solution offered to the puzzle (transitivity) simply denies one of the elements from which the puzzle arose. Of course, the step from φ to χ is “elementary” in the sense that, at least as far as consequence is concerned, it is no more mediated than those from φ to ψ and from ψ to χ . Indeed, in this sense, *every* valid step is an “elementary” step. The real trouble is then that the transitivity solution simply obliterates the *structure* presupposed by the puzzle, namely the distinction between elementary logical steps and non-elementary ones.

This suggests that the elementary/non-elementary distinction has been mislocated by the argument. It is not a distinction to be drawn at the level of *consequence*—rather, it is a distinction to be drawn at the level of *inference*. For present purposes, the notion of an elementary inference can be left at an intuitive level: it is understood in such a way that the inference of φ from ‘ φ and ψ ’ is elementary, whereas that of ‘ $[\varphi$ or $\chi]$ and ‘ $[\psi$ or $\nu]$ ’ from ‘ φ and ψ ’ is not. This is just as good, since the puzzle was an *epistemic* one and the drawing of an inference is one of the canonical ways in which the validity of an argument can be *recognized* by a subject. The transitivist herself has thus to acknowledge that the puzzle should properly be stated as the puzzle of how any inference can be substantially informative given that every elementary inference is least informative. Her own solution to the puzzle may then be revised as follows. *x-is-elementarily-inferable-from-y* is non-transitive, yet it entails *x-is-inferable-from-y*. Furthermore, the soundness of the inferences (with respect to consequence) makes it the case that *x-is-inferable-from-y* entails *x-is-a-consequence-of-y*. The transitivity of the latter in turn ensures that the final conclusion can be inferred from the initial premises, even if such an inference is non-elementary (and non-obvious).

The revised argument is coherent and might seem to offer a satisfactory explanation of the revised puzzle from a transitivist perspective. However, if *x-is-inferable-from-y* is itself transitive, the appeal to the transitivity of consequence is now exposed as a superfluous detour. For a simpler argument can run as follows. *x-is-elementarily-inferable-from-y* is non-transitive, yet it entails *x-is-inferable-from-y*. The transitivity of the latter in turn ensures that the final conclusion can be inferred from the initial premises, even if such an inference is non-elementary (and non-obvious).

Now, as has been so far explicated, *x-is-inferable-from-y* is very general and need not be transitive from a non-transitivist point of view. Yet, any theoretical employment of logic by a being whose cognitive architecture resembles that of humans is likely to require a *systematization* of the inferences whose validity can be recognized without further analysis into a small set of syntactic rules which are at least sound (and possibly complete) with respect to consequence. It is simply a fact that such a systematization provides the most effective method a human (and anyone with a similar cognitive architecture) can employ in order to explore what follows from what (just as something like the standard rules for addition, subtraction, multiplication and division together with something like decimal notation provides the most effective method a human (and anyone with a similar cognitive architecture) can employ in order to explore what gives what).

To stress, I don't think that there is an immediate connection between inference and syntactic rules: I think that it's clear on reflection that one can recognize the validity of an inference (and so endorse it) simply in virtue of one's appreciation of the logical concepts involved in it, even if one is not in possession of a set of syntactic rules which would allow the relevant derivation. Yet, as I have just noted, given the way human cognition works, there is every theoretical reason for transitivists and non-transitivists alike to accept a systematization of our pre-theoretical judgements of validity in terms of syntactic rules. And while there is no conceptual bar to the deductive system thus generated being itself non-transitive, it is of course a

desirable epistemic feature also from the non-transitivist perspective that the deductive system used in the study of a non-transitive consequence relation be itself transitive (in the sense that any application of a rule preserves the property of being a correct deduction). This feature is desirable because non-transitive consequence can be just as non-obvious as transitive consequence, whence the need arises for deductive techniques offering the epistemic gain flowing from the asymmetry between the non-transitivity of *x-is-obviously-deducible-from-y* and the transitivity of *x-is-deducible-from-y* (for an example of non-transitive logics with transitive sound and complete deductive systems see Zardini [2007b]). No further asymmetry between the non-transitivity of *x-is-an-obvious-consequence-of-y* and the alleged transitivity of *x-is-a-consequence-of-y* is required.

5.3.2 Consequence and Truth Preservation

We have already encountered in section 5.2.1 the (rather vague) claim that consequence is all about truth preservation, in the sense that whether the consequence relation holds between certain premises and conclusions is *wholly* determined by whether the conclusions preserve the truth of the premises. On the basis of an intuitive understanding of such notion of truth preservation, the following objection can be mounted.

The non-transitivist, we may assume, claims that, for some φ , ψ and χ , ψ follows from φ and χ from ψ , but χ does not follow from φ . So, since consequence requires truth preservation, the non-transitivist should concede that ψ preserves the truth of φ and that χ preserves the truth of ψ . Yet, on the face of it, truth preservation is a transitive relation: if ψ preserves the truth of φ and χ preserves the truth of ψ , then it would seem that χ also preserves the truth of φ . For suppose that ψ preserves the truth of φ and that χ preserves the truth of ψ , and suppose that φ is true. Then, since ψ preserves the truth of φ , ψ should be true as well. But χ preserves the truth of ψ , and so, since ψ is true, χ should be true as well. Thus, under

the supposition that φ is true (and that ψ preserves the truth of φ and that χ preserves the truth of ψ), we can infer that χ is true. Discharging that supposition, we can then infer (still under the suppositions that ψ preserves the truth of φ and that χ preserves the truth of ψ) that, if φ is true, so is χ , which might seem sufficient for χ 's preserving φ 's truth. The objection is completed by noting that truth preservation suffices for consequence, so that, since χ preserves the truth of φ , χ follows from φ , contrary to what the non-transitivist claims.

A proper assessment of the objection requires an adequate explication of the underlying notion of truth preservation. Picking up some threads of section 2.2.2, I know of no better way of spelling out this notion than in terms of a *conditional* statement: certain conclusions preserve the truth of certain premises iff, *if* every premise is true, *then* some conclusion is true (where, for the purposes of the discussion to follow, we can afford to remain rather neutral as to the exact behaviour of the conditional). Of course, so stated in terms of an unadorned indicative conditional, truth preservation alone does not *suffice* for consequence. Moreover, as I argue in Zardini [2007c], it is very doubtful that any sufficient non-circular strengthening of the intensional force of the conditional is available, and the very idea of the sufficiency of truth preservation risks making unintelligible some first-order debates about valid reasoning. Unfortunately, I have no space to rehearse those arguments here, and so I will simply have to assume their conclusion that there is no interesting sense of ‘truth preservation’ in which truth preservation suffices for consequence. Given this, I will also officially adopt the above explication of the notion (save for briefly considering a possible alternative at the end of this section).

It must be stressed that, if truth preservation does not suffice for consequence, the simple argument against the non-transitivist presented at the outset of this section results unsound. In particular, the argument breaks down at the step ‘since χ preserves the truth of φ , χ follows from φ ’, which assumes truth preservation to suffice for consequence. Yet, a more sophis-

ticated version of the argument could still be made to run against some applications that the non-transitivist envisions for her logics (for example, to the sorites paradox). Let us say that:

- An application of non-transitivity which only requires a verdict of invalidity concerning a transitivistically valid target argument is *weak*;
- An application which in addition requires the rejection of a commitment to accepting the conclusions of a transitivistically valid target argument all of whose premises are accepted⁹ is *intermediate*;
- An application which in addition requires the falsity of all the conclusions of a transitivistically valid target argument all of whose premises are true is *strong*.

Focus then on strong applications of non-transitivity. For some of these, the non-transitivist wishes to maintain that φ is true but χ false, even though ψ follows from φ and χ from ψ .¹⁰ However, the transitivity of truth preservation and its necessity for consequence seem to be sufficient to establish that, if

⁹For simplicity's sake, I will henceforth take sentences as the objects of acceptance and rejection. The whole discussion may be recast, more clumsily, in terms of acceptance of propositions.

¹⁰In fact, at least one of the most prominent such applications, i.e. the one to the sorites paradox, presents some recalcitrance against being fit into this mould. For that is an application with *multiple-premise* arguments, so that transitivity of truth preservation is not applicable to such chains of arguments in the direct way exploited by the revised argument in the text. One way to recover such an application would be to modify the relevant arguments, so as to bring all the premises $\varphi_0, \varphi_1, \varphi_2 \dots \varphi_i$ used at some point or other in the chain up front, collect them together in a single, long “conjunction” and carry them over from conclusion to conclusion adding to the relevant conclusion ψ the “conjuncts” $\varphi_0, \varphi_1, \varphi_2 \dots \varphi_i$ (scare quotes being used here since, for example in tolerant logics, standard conjunction has not the right properties to do the job and a new operation would have to be introduced instead). An alternative way of recovering the application would be to use a stepwise application of the alleged transitivity of truth preservation on the original arguments. Given the relevant chain of arguments $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_j$, with a first multiple-premise argument \mathbf{a}_i after \mathbf{a}_0 , transitivity of truth preservation can be used

φ is true, so is χ —a conditional which sits badly with the other commitments (truth of φ , falsity of χ) the non-transitivist would wish to undertake (exactly how badly it sits will depend of course on the details of the logic—for example, these claims are jointly inconsistent in all the various classical tolerant logics of chapter 4). Unfortunately, the series of considerations concerning context dependence, the semantic paradoxes and higher-order indeterminacy developed in Zardini [2007c] would appear to show that even the *necessity* of truth preservation for consequence cannot be sustained in full generality.

Before seeing how the argument from truth preservation may still be made to run in the presence of these limitations, I would like to undertake a brief digression on truth preservation which I hope will prove instructive for understanding some features of non-transitive logics. For it may be replied to the foregoing that consequence is trivially guaranteed to preserve at least *some kind* of truth, even if not truth *simpliciter*. For consequence is often defined as preservation, for every model \mathfrak{M} , of truth in \mathfrak{M} . Consequence would then be trivially guaranteed to preserve truth in a model. In a sense, already the identification of consequence with the *preservation* of something or other (in some structure or other) just begs the question against the non-transitivist, as there are no *compelling* intuitive or theoretical grounds for such an identification and one can define well-behaved consequence relations without appealing to anything recognizable as preservation of something or other (in a some structure or other).¹¹

to yield the truth of at least one (possibly all) of the relevant premises of \mathfrak{a}_i . Having so secured the truth of those premises, the truth of any other premise (possibly none) of \mathfrak{a}_i is conceded by the non-transitivist (for example, in the case of the sorites paradox, this would be the truth of the relevant instance of (3)). The procedure can be applied again until the truth of the conclusion of \mathfrak{a}_j is reached.

¹¹Even granting an identification of consequence with truth preservation in some structure or other, it is well-known that the standard set-theoretic notion of model, which replaces the generic notion of structure, has serious drawbacks in the analysis of the consequence relation of expressively rich languages. I won't go here into this further aspect of the complex relation between consequence and preservation of truth in a structure (see McGee [1992] for a good introduction to some of these issues).

Tolerant logics are an example of such a consequence relation developed in a non-transitive framework; an example driven by a completely different kind of consideration (not affecting transitivity) can be found in Martin and Meyer [1982]. The general structural point in these logics is that the collection of designated values relevant for the premises is not identical with the collection of designated values relevant for the conclusions: in the tolerant logics of chapter 4 the former is in effect a proper subcollection of the latter, whereas the logic **S** of Martin and Meyer [1982] can be seen as a form of the dual position, in which the former is a proper supercollection of the latter (thereby leading to the failure not of transitivity, but of another property considered by Tarski [1930] essential for a consequence relation—namely, reflexivity).¹² Let me be clear. It is not the case that a representation in terms of non-identity of the two collections suffices to ensure the non-transitivity of a logic (see Smith [2004] for an example of a transitive logic generated by such a representation). Nor is it the case that a representation in terms of non-identity of the two collections is essential for every non-transitive logic (see Smiley [1959] for an example of a non-transitive logic where no such natural representation seems to be forthcoming). Yet such a representation, where available, is a fruitful point of entry to at least one of the key thoughts behind the logic and as such has been (in section 4.2.2) and will be deployed (in section 5.4.4) in this essay. The representation also connects up neatly with usual representations of transitive logics (preservation of designated value in some structure or other), showing how non-transitivity arises from a very natural and straightforward generalization of the usual model-theoretic representation of a (transitive) consequence relation.

Coming back to our main line of thought and in spite of the qualms that the foregoing considerations should warrant against any straightforward appeal to truth preservation, I am willing to concede that, in the cases under discussion, consequence does indeed require truth preservation. However, I

¹²Thanks to Bob Meyer for a very helpful discussion of **S** and to Graham Priest for pointing out to me the duality connection.

reject that the non-transitivist is thereby committed to χ 's being true if φ is. For this conclusion (and so the transitivity of truth preservation) follows from the premises that ψ preserves the truth of φ and that χ preserves the truth of ψ only under the assumption that the consequence relation of the metalanguage (the language in which we talk about the truth of φ , ψ and χ) is transitive (as against the non-transitivity of the consequence relation of the object language in which we talk about whatever φ , ψ and χ talk about). More explicitly: under the current interpretation of truth preservation, the validity of the argument to the conclusion that χ is true if φ is (and so to the transitivity of truth preservation) boils down to the validity of the argument form 'If P_0 , then P_1 ; if P_1 , then P_2 . Therefore, if P_0 , then P_2 ', which should however be invalid in a non-transitive logic (it is for example invalid in all tolerant logics of chapter 4).

Moreover, no deviance from classical logic worthy of this name should grant the assumption that the consequence relation of the metalanguage \mathcal{M} talking about the truth and falsity of the sentences of the object language \mathcal{O} over which a deviation from classical logic is envisaged should itself be classical. For, given any decent theory of truth, this will be sufficient to reintroduce classical logic in \mathcal{O} itself. Consider for example a deviant intuitionist logician. Were she to accept 'Either ' φ ' is true or ' φ ' is not true' in \mathcal{M} (for φ belonging to \mathcal{O}), the most natural theory of truth (namely one such that φ 's being not true implies that φ is false and so that 'It is not the case that φ is true' would commit her to 'Either φ or it is not the case that φ '.

The point need not exploit the full equivalence between φ and ' φ is true' which is induced by the *enquotation/disquotation* schema:

(ED) P iff ' P ' is F

(here and in what follows, make a proviso for the semantic paradoxes if you like). We can produce similar results only e.g. with the right-to-left direction of (ED). If a metalinguistic necessity predicate were to behave classically,

the intuitionist would have to accept in \mathcal{M} ‘Either ‘ φ ’ is necessary or ‘ φ ’ is not necessary’ (for φ belonging to \mathcal{O}). Given that ‘‘ φ ’ is not necessary’ is classically equivalent with ‘‘It is not the case that φ ’ is possible’’, she would have to accept ‘Either ‘ φ ’ is necessary or ‘It is not the case that φ ’ is possible’’, whence, by substituting ‘Either φ or it is not the case that φ ’ for ‘ φ ’, she would have to accept ‘Either ‘Either φ or it is not the case that φ ’ is necessary or ‘It is not the case that either φ or it is not the case that φ ’ is possible’’, which entails in (any suitable modal extension of) intuitionist logic ‘‘Either φ or it is not the case that φ ’ is necessary’. By right-to-left (ED), she would have to accept ‘Either φ or it is not the case that φ ’.¹³ Indeed, up to the very last step, the previous argument would go through also for necessity-like metalinguistic predicates¹⁴ which fail to satisfy either direction of (ED), such as ‘justified’. Thus, if a metalinguistic justification predicate were to behave classically, the intuitionist would have to accept in \mathcal{M} ‘‘Either φ or it is not the case that φ ’ is justified’ (for φ belonging to \mathcal{O}).

More generally, the point will hold for any two languages \mathcal{L}_0 and \mathcal{L}_1 as soon as some of the notions expressible in \mathcal{L}_1 are best thought of as exhibiting the same problematic properties which motivate a revision of the

¹³This argument should make clear what I mean when I say that a metalinguistic predicate Φ “behaves classically”. What I mean is that every instance of an argument form valid in classical first-order logic in which Φ occurs essentially is valid (treating quotation names as constants) and, in addition, that Φ has a dual Ψ satisfying the schema ‘‘ P ’ is not Φ iff ‘It is not the case that P ’ is Ψ ’. In particular, note that since ‘ a is F ’ is not an argument form valid in classical first-order logic, we cannot straightforwardly assume ‘‘Either φ or it is not the case that φ ’ is necessary’ (maybe on the grounds that ‘Either φ or it is not the case that φ ’ is classically valid), whose most fine-grained form is precisely ‘ a is F ’. Only the object-language logic—which in our example is intuitionist—is allowed, as it were, to “look inside” quotation names, whence it is crucial that, in our argument, ‘‘It is not the case that either φ or it is not the case that φ ’ is not possible’ is valid in (any suitable modal extension of) intuitionist logic.

¹⁴A metalinguistic predicate Φ is *necessity-like* iff ‘‘ φ ’ is Φ ’ and ‘‘ ψ ’ is Φ ’ entail ‘‘ φ and ψ ’ is Φ ’ and φ entails ‘‘ φ ’ is Φ ’ if φ is the trivial truth (that is, the truth entailed by anything).

logic for \mathcal{L}_0 (maybe exactly because of some systematic connection that these notions bear to notions expressible in \mathcal{L}_0 , as is the case for the notion of truth when \mathcal{L}_1 is the metalanguage of \mathcal{L}_0 and truth is governed by (ED)). This more general point fatally affects also a revision of the argument from truth preservation which replaces the notion of truth preservation with that (logically identical) of closure of knowledge (or of other epistemic properties) under logical consequence.¹⁵

It is worth seeing in closing how the argument from truth preservation fares if the current (and standard) explication of the notion of truth preservation in terms of the notion of conditionality is rejected in favour of a *primitive* relation of truth preservation. While under the former explication there was a logical guarantee that the relation is transitive (at least assuming a transitive logic!), now that guarantee is lost and the alleged transitivity must be postulated as a specific law governing the relation. Clearly, there is no need on the non-transitivist's part to accept this postulation. What is important to note now is that even the acceptance of a transitivity postulate for the relation would not by itself wreck havoc for a strong application of non-transitivity like the one to the sorites paradox, for what is really at issue there is not so much simple transitivity, but the stronger assumption that a finite chain of elements connected by a relation R is such that its first element bears R to its last element. This is in effect the finite chain-transitivity assumption (CTRANS ^{R^f}), which, in a non-transitive logic, is usually stronger than the transitivity assumption (TRANS ^{R}) (see section 4.5.6). Even though there is little need for the non-transitivist to pursue this latter strategy in the case of a primitive truth-preservation relation, I think there is good reason for her to embrace it in other cases where independent grounds support the transitivity of a particular relation, as in the case of the identity relation (see again section 4.5.6).

¹⁵Thanks to Stephen Schiffer for interesting discussions on this kind of argument.

5.4 The Non-Transitivist's Picture

5.4.1 Non-Logical/Logical Dualism and Non-Transitivism

We have only started to scratch the surface of the philosophical underpinnings of a non-transitive logic. The rebuttal of the two previous objections has helped to dispel some misunderstandings of what non-transitivity of consequence amounts to, but not much has yet been offered in the way of a positive characterization of a conception of consequence as non-transitive.

I think more progress on this issue can be made by asking the question as to why consequence is usually assumed to be transitive. Consider the following natural picture (where, for simplicity's sake, we take sets to be the terms of the consequence relation and restrict our attention to single-conclusion arguments). The laws of logic can be seen as an *operation*¹⁶ which, applied to a set of facts X , yields another such set Y (possibly identical with X). They apply with necessity to X , yielding Y , but they also apply with

¹⁶ The study of consequence as an operation rather than relation goes back at least as far as Tarski [1930] (see Wójcicki [1988] for a recent comprehensive study within this approach). Under the simplifying assumptions made at the start of this section, the properties of reflexivity, monotonicity and transitivity of a consequence relation correspond to the following properties of a consequence operation cons from sets of sentences to sets of sentences:

- (i) $X \subseteq \text{cons}(X)$ (increment);
- (ii) If $X \subseteq Y$, $\text{cons}(X) \subseteq \text{cons}(Y)$ (monotonicity);
- (iii) $\text{cons}(\text{cons}(X) \cup Y) \subseteq \text{cons}(X \cup Y)$ (union-adjoint subidempotency).

An operation satisfying conditions (i)–(iii) is a *Tarski closure operation*. A generalization of closure operations for modelling multiple-conclusion consequence relations are *Scott closure operations* (see Scott [1974]). Further generalizations dealing with collections more fine-grained than sets are possible (see Avron [1991] for a start). Closure operations have first been identified by Kuratowski [1922].

necessity to Y , yielding another set Z (possibly identical with Y). It may therefore seem to follow from this natural picture that X is already sufficient to yield with logical necessity Z —i.e. that the laws of logic applied to X already yield Z . This would mean that the result (Z) of the operation of the laws of logic on the result (Y) of the operation of the laws of logic on X is included in the result (Y) of the operation of the laws of logic on X —the operation would be subidempotent,¹⁷ thus validating a form of transitivity for the corresponding relation of consequence. However, as in the case of the objection from truth preservation, the conclusion that X yields with logical necessity Z (that is, that, if all the members of X hold, so do all the members of Z), in its characteristic obliteration of the logical role played by the intermediary conclusion (that is, that all the members of Y hold), implicitly relies on transitivity. Given that facts-talk is no less non-neutral than truth-talk, the non-transitivist should not be seen as committed to the denial of the natural picture.

This point is crucial. It is very tempting to try to make sense of the non-transitivist's position, especially in its intermediate and strong applications,¹⁸ as relying on a distinction between two different kinds of facts, the non-logical and the logical. *Non-logical facts* are those provided, as it were, by the world itself, such as the fact that snow is white; the fact that, if snow is white, it reflects light; the fact that every piece of snow is white. *Logical facts* are those that hold in virtue of the application of the laws of logic to the non-logical facts, such as the fact that either snow is white or grass is blue (holding in virtue of the application of the laws of logic to the fact that snow

¹⁷An operation op is *subidempotent* on a set X and ordering \leq iff, for every $x \in X$, $\text{op}(\text{op}(x)) \leq \text{op}(x)$. Under the simplifying assumptions made at the start of this section, in our case \leq is simply subset inclusion. It's easy to check that, with conditions (i) and (ii) of fn 16 in place, subidempotency on subset inclusion implies union-adjoint subidempotency.

¹⁸Henceforth, I will mainly talk about non-transitive logics tailored to intermediate and strong applications of non-transitivity, as these are arguably the philosophically most interesting cases to be made sense of. This qualification must be understood as implicit in the following.

is white); the fact that snow is white and grass is green (holding in virtue of the application of the laws of logic to the fact that snow is white and the fact that grass is green); the fact that something is white (holding in virtue of the application of the laws of logic to the fact that snow is white). Of course, much more would have to be said about how to draw exactly the non-logical/logical distinction, but I take it that we have an intuitive grasp of it (as witnessed by our intuitive judgements in the foregoing cases) which will be sufficient for present purposes.

The non-transitivist would then be seen as rejecting that the application of the laws of logic to the logical facts always results in non-vacuous effects. By the application of the laws of logic “having non-vacuous effects” I mean that they manage to force the fact described by the conclusion to hold whenever every fact described by the premises holds (where, as in the case of truth preservation, this is to be spelled out in terms of a conditional statement). Whenever at least one of the facts of the set to which an application of the laws of logic is envisaged is logical, it would then be open to the non-transitivist to deny non-vacuous effects to the application of the laws of logic to that set. To stress, according to such *non-logical/logical dualism*, having non-vacuous effects of application does not coincide with being valid: an argument form may be universally valid but, when applied to a set of facts which correspond to its premises and some of whose members are logical, fail to force the fact corresponding to its conclusion to hold. Thus, on this dualist view, the validity of an argument is one thing and a different, stronger thing is the argument's being able to force the fact described by its conclusion to hold whenever every fact described by its premises holds.

There might be an important distinction between accepting an inference and accepting the corresponding conditional, in the sense that one can accept the former while rejecting the latter (see e.g. Field [2006]; Zardini [2007c]). As should be clear, I am understanding the dualist as someone who, in our original case, not only rejects the conditional that, if all the members of *Y* hold, so do all the members of *Z*, but also refuses to infer the conclusion

that all the members of Z hold from the premise that all the members of Y hold. As opposed to her rejection of the conditional, her rejection of the inference is what she shares with a non-dualist non-transitivist. Still, I want to understand her rejection of the inference as *grounded* in her rejection of the conditional. Her rejection of the conditional must then be understood in a suitably strong sense, as implying some kind of possibility of the antecedent's holding and the consequent's failing.

Under a certain extremely plausible assumption, this dualist position does in effect require restrictions on the transitivity of consequence, at least if failure to yield non-vacuous effects is to be allowed when an argument is applied to a set of logically contingent facts. For, whenever a restriction on the non-vacuous effects of the application of the laws of logic to some non-empty set of logically contingent facts described by the premises $\varphi_0, \varphi_1, \varphi_2 \dots \varphi_i$ is envisaged (so that the inference to the conclusion ψ , which follows from them, is rejected), the j th premise must itself be in the transitive closure of the consequence relation to some non-empty set Γ_j whose members describe non-logical facts. This is so because, even if not every logically contingent fact is non-logical (consider for example the fact that something is white), it is extremely plausible to assume that every logically contingent fact is ultimately grounded in a non-logical fact in such a way as to follow from it (here I won't try though to justify this assumption). If transitivity were then to hold unrestrictedly, ψ would already follow from $\Gamma_0 \cup \Gamma_1 \cup \Gamma_2 \dots \cup \Gamma_i$. Since however the members of this set all describe non-logical facts, there would be no dualist bar to the application of the laws of logic having non-vacuous effects, and so the fact described by ψ would be forced to hold as well.

Interesting as such a dualist position may be, it is crucial to see that a non-transitivist is not committed to it—indeed, that versions of non-transitivism are committed to the negation of its tenet that, to take a single-premise case, χ might follow from ψ even if it is not the case that, if the fact described by ψ holds, so does the fact described by χ (this might be so when the fact described by ψ in turn only holds because ψ follows from another sentence φ

describing a non-logical fact that holds). For a non-transitivist may well hold, against the dualist, that the laws of logic apply with non-vacuous effects to whichever facts turn out to hold: such a non-transitivist would accept that, if the fact described by ψ holds, no matter on what grounds, so does the fact described by χ .

Switching from facts-talk to truth-talk, in the situation envisaged the dualist denies that, even though χ follows from ψ , it must be the case that, if ψ is true, so is χ . The non-dualist non-transitivist, on the contrary, accepts that. This should however not obscure a deeper point of agreement between the two positions. For reflect that, in a non-transitive framework, the dualist can be seen as doing one thing by means of a quite different one. That is, to come back to the last example, she can be seen as rejecting a commitment to a consequence (χ) of what she is logically committed to (ψ) by maintaining that it is not the case that, if ψ is true, so is χ (she is logically committed to ψ because it follows from φ , to which she is committed on non-logical grounds). Even if a non-transitivist can disagree with the latter, she cannot but agree with the former, since the rejection of being committed to a consequence of what she is logically committed to is arguably the *crux* of her disagreement with the transitivist, the place at which their different metalinguistic judgements about validity are finally reflected in a clash of object-linguistic attitudes (in this case, the transitivist's acceptance of χ against the non-transitivist's non-acceptance of χ). A non-transitivist need not disagree with the transitivist as to whether, if ψ is true, so is χ , but she has to disagree with the transitivist's willingness to undertake a commitment to χ 's being true on the sole basis of her logical commitment to ψ 's being true.

It should by now be clear that—in opposition to the dualist—a non-dualist non-transitivist can legitimately insist on not explaining her position in terms of some deviant conception of the relation between the truth of ψ and the truth of χ (such that it is not the case that, if ψ is true, so is χ), just as someone who rejects the law of excluded middle need not explain

her position in terms of a deviant (e.g. gappist) conception of the relation between the truth of a sentence and the truth of its negation (e.g. such that it might be the case that neither a sentence nor its negation are true), or just as someone who rejects disjunctive syllogism need not explain her position in terms of a deviant (e.g. dialetheist) conception of the relation between the truth of a sentence and the truth of its negation (e.g. such that it might be the case that both a sentence and its negation are true). As in all those other cases, the heart of the logical revisionism being proposed is a certain conception of what counts as a *correct pattern of reasoning* rather than a certain deviant conception of what truth is. The thought that truth, its properties and laws is what consequence is all about (see e.g. Frege [1893], pp. XV–XVI) only risks making unintelligible the point of these proposals.

5.4.2 Logical Nihilism and Non-Transitivity

Still, regarded now as a proposal as to what counts as a correct pattern of reasoning, non-transitivity may look perilously close to a *logical nihilism* which rejects the universal validity of all the argument forms that are the last coordinate of a sequence lacking the left-transitivity property (of course, even in a non-transitive logic not every argument form is usually such). For take without loss of generality single-conclusion arguments $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_i$ such that $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_{i-1}$ are of forms $\mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \mathbb{F}_{i-1}$ and \mathbf{a}_i of form \mathbb{F}_i , and such that the relevant instance of (T^l) fails for them. Then $\langle \mathbb{F}_0, \mathbb{F}_1, \mathbb{F}_2 \dots \mathbb{F}_i \rangle$ lacks the left-transitivity property. Consider then a non-transitivist who accepts on non-logical grounds all the premises Γ of $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_{i-1}$ and accepts (on non-logical or logical grounds) the premises Λ of \mathbf{a}_i which are not conclusions of any of $\mathbf{a}_0, \mathbf{a}_1, \mathbf{a}_2 \dots \mathbf{a}_{i-1}$. Such a non-transitivist would then have to accept all the premises Λ, Θ of \mathbf{a}_i . Yet, given the foregoing explanation of what the non-transitivist regards as a correct pattern of reasoning, she need not regard herself as committed to the conclusions Ξ of \mathbf{a}_i . How could she then still maintain that \mathbf{a}_i is *valid*, given that she accepts all its premises but refuses

to infer its conclusions? Is her refusal to infer the conclusions not an implicit admission that she does not regard \mathbf{a}_i as valid (and thus that she does not regard \mathbb{F}_i , which is instantiated by \mathbf{a}_i , as a universally valid argument form)? Note how these questions are particularly pressing for a non-dualist non-transitivist, since she cannot help herself to the dualist doctrine that it is not the case that, if every premise of \mathbf{a}_i is true, so is some conclusion of \mathbf{a}_i , doctrine which would certainly go some way towards explaining the refusal to infer the conclusions of \mathbf{a}_i .

Before addressing these urgent questions, note that nihilism, as against dualism, is not a possible option for a non-transitivist, at least in the following sense. It might well be that all of the nihilist, the dualist and the non-dualist non-transitivist accept all the premises Γ of a classically valid argument \mathbf{a} of form \mathbb{F} while refusing to infer its conclusions Δ . The dualist might do this because, even though she recognizes \mathbf{a} as valid, she regards some coordinate of Γ as describing a logical fact, and so she rejects that, if every coordinate of Γ is true, so is some coordinate of Δ (and refuses to infer Δ from Γ). The non-dualist non-transitivist might do this because, even though she recognizes \mathbf{a} as valid and accepts that, if every coordinate of Γ is true, so is some coordinate of Δ , she still refuses to infer Δ from Γ (on grounds which we will explore shortly). The nihilist, however, refuses to infer Δ on the very simple grounds that she does not regard \mathbb{F} as a universally valid argument form and that, in particular, she regards its instance \mathbf{a} as invalid. She thus cannot regard \mathbf{a} as involved in a possible failure of the transitivity of consequence, as the dualist and the non-dualist non-transitivist do. Nihilism is an *alternative* to non-transitivism, not one of its species.

To come back to the question as to how a (non-dualist) non-transitivist can recognize as valid an argument whose premises she accepts and whose conclusions she does not accept, we must observe that the connection, presupposed by this question, between recognizing an argument as valid and inferring the conclusions if one also accepts its premises is much less straightforward than it might seem at first glance. One can recognize an argument

as valid and accept its premises while still not inferring its conclusions because one is somehow prevented by external circumstances from doing so (by a threat, a psychological breakdown, a sudden death etc.). Or because one fails to recognize, maybe on account of their syntactic complexity, that the premises (conclusions) are indeed premises (conclusions) of an argument one recognizes as valid. Or because one has a general policy of not inferring conclusions, maybe for the reason that one has been told by one's guru that every inference is sacrilegious. Or maybe because one simply cannot be required to infer all the conclusions of all the arguments one recognizes as valid and whose premises one accepts—this is certainly not a requirement on resource-bounded rationality, and, unless each and every single truth is an aim of belief, it is not clear why it should even be a requirement on resource-unbounded rationality.

The previous counterexamples may seem to trade on “*deviant*” cases. Still, unless a plausible independent characterization of “deviancy” is provided (a highly non-trivial task), the objection against the non-transitivist would seem to lose much of its force—why should non-transitivism be itself classified as one of the “deviant” cases (which we know from the foregoing counterexamples generally to exist)? Be that as it may, stronger, only slightly more controversial counterexamples can be given where there is actually *epistemic force against the inference's being drawn*:

VANN Vann may be told by a source he is justified to trust that, if Vann's initial is ‘V’, then Vann is a horribly bad *modus-ponens* inferrer (which does not imply that Vann is horribly bad at recognizing the validity of *modus-ponens* arguments). Vann may also know that his initial is ‘V’. In these conditions, there would plausibly be epistemic force against Vann's inferring the conclusion that he is a horribly bad *modus-ponens* inferrer, even though Vann may recognize the validity of the relevant instance of *modus ponens*—plausibly, it would be unwarranted to believe that one is a horribly bad *modus-ponens* inferrer exactly *via*

a *modus-ponens* inference.¹⁹

DAVID Sincere and modest David might believe of each *of* the 1,000 substantial statements of his new history book that *that* statement is true—if sincere David didn't really believe a statement to be true, why would he have put it in the book in the first place? Together with the certainly true assumption that those are all the statements in his book, it follows that all the statements in David's book are true. David recognizes the argument as valid, yet there is epistemic force against modest David's inferring the conclusion that all the statements in his book are true.²⁰

¹⁹Note in passing that this counterexample can easily be turned into a counterexample against the *prima facie* very plausible closure principle that “knowing p_1, \dots, p_n , competently deducing q , and thereby coming to believe q is in general a way of coming to know q ” (Williamson [2000b], p. 117; cf Hawthorne [2004], p. 33). Thanks to Daniele Sgaravatti, Martin Smith and Crispin Wright for discussion of this and similar examples.

²⁰This counterexample is of course a version of the preface paradox (see Makinson [1965] and Christensen [2004] for a recent congenial discussion). The counterexample is even more telling given the relationships between probabilistic reasoning and some kinds of non-transitive deductive consequence relations remarked upon in fn 7. As Crispin Wright has emphasized to me in conversation, there is an important asymmetry between the counterexample, which crucially relies on the fact that there is epistemic force against accepting the conjunction of all the premises, and some application of non-transitivity, where one would wish not just to accept all the premises, but also their conjunction (more or less equivalently, one would wish not only to accept of every premise that it is true, but also to accept that every premise is true). (Indeed, this is why in sections 4.5.7, 4.5.8 I went to some length to show that **CIIT**¹ and **NDCIIT**¹ have the nice feature that the consistency in them of the (finite) axiomatic base of a theory \mathcal{T} implies the consistency in them of the conjunction of the axioms of \mathcal{T} .) The point of the counterexample is however only to show that there can be epistemic force against an inference's being drawn even if the argument is recognized as valid and all its premises are accepted. The *source* of this epistemic force in **DAVID** is such that it only applies to multi-premise arguments, whereas the source of the force in some application of non-transitivity will evidently not carry this restriction (Harman [1986], pp. 11–20 is the *locus classicus* for the problematization of the connection between validity and inference).

These counterexamples are certainly sufficient to open up conceptual space for a genuinely non-transitivist (rather than logical nihilist) position. Still, more needs to be said by way of a positive explanation of the refusal to draw an inference when an argument is recognized as valid and all its premises are accepted (at least, more needs to be said if one wants to avoid giving the explanation that the dualist gives). Moreover, a doubt now arises as to the very point of non-transitivism: if a story needs to be told anyway in order to vindicate the rationality of accepting all the premises of an argument recognized as valid while refusing to infer its conclusions, could such a story not be applied in a transitive framework, recognizing, say, that, since, [for every $\varphi \in \text{ran}(\Theta)$, Γ entails Δ, φ and Λ, Θ entails Ξ], Λ, Γ does entail Δ, Ξ , and accepting Λ, Γ , while refusing to infer Δ, Ξ ? Such a doubt would certainly be pressing at least for intermediate and strong applications of non-transitivity.

The situation which seems to be emerging is this. The non-transitivist needs to differentiate between the acceptance-related normative property of being N_0 (which triggers the normative force of consequence once all the *premises* of an argument recognized as valid are N_0) and the acceptance-related normative property of being N_1 (which consequence so generates with respect to the *conclusions* of the argument), in such a way that being N_1 need not in turn imply being N_0 (and so in such a way that the acceptance-related normative property of being N_0 need not be closed under logical consequence). Such a distinction between acceptance-related normative properties would allow the non-transitivist to insist that, in the cases where she accepts all the premises of an argument recognized as valid while refusing to infer its conclusions, this is so because the premises are only N_1 and not N_0 . It would also allow her to reply to the awkward question of the last paragraph by saying that, since Λ, Γ are N_0 , if consequence were transitive, Δ, Ξ would have to be N_1 and so she would after all be committed to accepting it (since *x-is- N_1* would arguably have to be at least so strong as to imply *one-is-committed-to-accepting-x*). To the identification of

the acceptance-related normative properties of being N_0 and of being N_1 we must now turn.

5.4.3 The Normativity of Consequence

Let us go back without loss of generality to the situation where the non-transitivist accepts φ for non-deductive reasons,²¹ accepts both that ψ follows from φ and that χ follows from ψ but does not accept that χ follows from φ . I say that, in such a situation, her non-transitivism is sufficient to save her from a commitment to accepting χ . The simple logical point that, in a non-transitive logic, it need not follow that, if φ is true, so is χ , has already been expounded in the rebuttal of the objection from truth preservation. It is now time to see how the non-transitivist can escape *in general* a commitment to accepting χ , whether *via* the previous conditional or any other route. To see this consider that, quite plausibly, one has deductive reasons²² to accept no more than the consequences of what one has *non-deductive* reasons to accept,²³ and our non-transitivist only has non-deductive reasons to accept φ . Since ψ is indeed a consequence of φ , and since she has non-deductive reasons to accept φ , she has indeed deductive reasons to accept ψ , and so she is indeed committed to accepting ψ .²⁴ However, χ need *not* be a consequence

²¹Henceforth, by ‘having non-deductive reasons to accept φ ’ and its like I really just mean ‘having reason to accept or accepting φ not because conclusion of a deductively valid argument all of whose premises one has reason to accept or accepts’—acceptance for non-deductive “reasons” need not be objectively well-grounded at all.

²²Henceforth, by ‘having deductive reasons to accept φ ’ and its like I really just mean ‘having reason to accept φ because conclusion of a deductively valid argument all of whose premises one has non-deductive reasons to accept’ (see fn 21).

²³Note that this principle can quite generally be held to exhaust the normative force of consequence (at least as far as the aspects we are concerned with here go), no matter whether the underlying consequence relation is transitive or non-transitive (see below in the text).

²⁴This last step should make clear that, to simplify the discussion, I am assuming that the relevant (non-deductive or deductive) reasons to accept a certain sentence are always so strong as to imply a commitment to accepting that sentence.

of φ (at least, if consequence is non-transitive!), and so she needs not be committed to χ simply because she has non-deductive reasons to accept φ (even though she is committed to ψ and has non-deductive reasons to accept that, if ψ is true, so is χ !).

Drawing on the foregoing implicit distinction between *having non-deductive reasons to accept* and *having deductive reasons to accept*, I suggest that we identify the acceptance-related normative properties of x -is- N_0 and x -is- N_1 with one-has-non-deductive-reasons-to-accept- x and one-has-deductive-reasons-to-accept- x respectively. I take it that this distinction has some very intuitive appeal and import in our evaluation of reasons: we would ordinarily distinguish between one's reasons to accept a certain sentence being so strong as to permit (or even mandate) acceptance of whatever follows from that sentence and one's reasons to accept a certain sentence being simply strong enough as to permit (or even mandate) acceptance of that sentence. In the former case, one's reasons allow (or even mandate) one to take the sentence as a starting point for further reasoning, whereas in the latter case they only allow (or even mandate) one to take the sentence as a terminal point of acceptance (see Smith [2004], pp. 196–9 for a defence of this distinction within a transitive framework). In the lights of the remarks already made in fn 7 and in connection to **DAVID**, it should go without saying that the distinction also makes perfectly good probabilistic sense.²⁵ Neither acceptance-related normative property implies actual acceptance, whilst being actually accepted implies both being N_0 (see fn 21) and being N_1 (from fn 21, (NDD) (see below) and reflexivity of consequence—indeed, by reflexivity of consequence and (NDD), being N_0 itself implies being N_1). This is so because both acceptance-related normative properties pertain to what *ought to be* the case rather than to what *is* the case.

²⁵Indeed, in the case of the strong application of non-transitivity to the sorites paradox, I actually think that the fact that, in a sense to be made more precise, vague belief exhibits probabilistic structure is the grain of truth in the thought (for which see Field [2000]; Field [2003b]; Schiffer [2003], pp. 178–237; Wright [2007a]) that vague belief is essentially a kind of partial belief.

In such a distinction between acceptance-related normative properties, we see how a trace of non-logical/logical dualism does necessarily remain in the non-dualist non-transitivist's position: only, the distinction is not between two different kinds of *facts*, but between two different kinds of *reasons for the acceptance of a sentence*—either non-deductive or deductive. In view of this distinction, the non-dualist non-transitivist can be seen not as endorsing the rather exotic restriction to non-logical facts of the non-vacuous effects of the application of the laws of logic, but as adhering unswervingly both to the verdicts of validity and invalidity issued by her non-transitive logic and to the general principle introduced two paragraphs back that the commitments generated by the laws of logic on a certain *position* (a collection—set, multi-set, sequence etc.—of sentences accepted for non-deductive reasons) coincide with the logical consequences of that position. That is, if one has non-deductive reasons to accept a certain collection of sentences, one is indeed committed by logic and those very same reasons to accepting each and every consequence of them, but also committed only to that (at least by logic and those very same reasons), so that, if logic is non-transitive, one is not committed by logic and those very same reasons to accepting a consequence of a consequence of one's position which is not already a consequence of one's position (such, as it were, “consequences at one remove” of course do not exist if logic is transitive, but they do if it isn't).

It is crucial to see that, as I have already indicated in fn 23, this basic principle governing consequence can be accepted *by non-transitivists and transitivists alike as exhausting its normative force* (at least as far as the aspects we are concerned with here go), and can be fleshed out as the following principle of *connection between having non-deductive reasons to accept and having deductive reasons to accept*:

- (NDD) Consequence only maps non-deductive reasons to accept a collection of sentences (the premises) onto deductive reasons to accept a collection of sentences (the conclusions).

Note that, *pending any further specification of the properties of the consequence relation*, (NDD) does not imply that consequence in turn maps deductive reasons to accept a collection of sentences onto deductive reasons to accept a collection of sentences. In other words, pending any further specification of the properties of the consequence relation, while (NDD) does imply that the normative force of consequence applies to premises accepted for non-deductive reasons, producing deductive reasons to accept (and hence commitments to accepting) the conclusions, it does not imply that such a force applies to premises one has simply deductive reasons to accept. Indeed, the non-transitivist can be seen as exploiting exactly the fact that, by itself, (NDD) does not imply the stronger principle of *preservation of deductive reasons to accept*:

(DD) Consequence maps deductive reasons to accept a collection of sentences (the premises) onto deductive reasons to accept a collection of sentences (the conclusions).

In particular, the non-transitivist can be seen as accepting (NDD) while rejecting (DD): on her view, one need not be committed to accepting consequences of commitments generated by logic.²⁶

Let me stress that the restriction operated by (NDD) (as against (DD)) of the normative force of consequence to premises accepted for non-deductive reasons will actually be far less draconian than it might at first appear to be for many interesting non-transitive consequence relations. Firstly, many such relations do enjoy unqualified transitivity properties (as defined in section 4.2.3) for a large number of argument forms. Restricted to such argument

²⁶(NDD) and (DD) are *closure principles*, in the sense that they say what one *has reasons to do* (accept or reject certain collections of sentences) given that one has reasons to do something (accept or reject certain collections of sentences). By contrast, (AR) in section 4.4.1 is a *coherence principle*, in the sense that it says what one *has reasons not to do* (accept or reject certain collections of sentences) given that one has reasons to do something (accept or reject certain collections of sentences).

forms, (NDD) will imply (DD). Secondly, even if an argument form does not enjoy an unqualified transitivity property, for many interesting intermediate and strong applications of non-transitivity the counterexamples will be circumscribable to a very specific class. Restricted to arguments outside of this class, (NDD) will still imply (DD) (where consequence will be material rather than formal). For example, in the case of the strong application of non-transitivity to the sorites paradox pursued in chapter 4, (NDD) will still imply (DD) restricted to the good classical arguments (see section 2.1). As I have already indicated, these can be roughly identified with those arguments which do not fallaciously exploit tolerance principles to go down the slippery slopes associated with vague predicates. Thus, in the case of the strong application of non-transitivity to the sorites paradox, the acceptance-related normative property triggering the normative force of consequence can be taken to be the weaker one-has-non-soritical-reasons-to-accept- x rather than the stronger one-has-non-deductive-reasons-to-accept- x .

It seems to me that, in her joint acceptance of (NDD) and rejection of (DD), the non-transitivist is occupying a reasonable position, given that the following theorem holds for every reflexive consequence relation:

Theorem 5.4.1. (NDD) *implies* (DD) *iff the consequence relation is transitive.*

Proof.

- *Left-to-right.* We prove the contrapositive and focus without loss of generality on failure of left transitivity. Take a reflexive non-transitive consequence relation \mathbf{L} such that, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash_{\mathbf{L}} \Delta, \varphi$ and $\Lambda, \Theta \vdash_{\mathbf{L}} \Xi$, but $\Lambda, \Gamma \not\vdash_{\mathbf{L}} \Delta, \Xi$, and consider an intermediate application of \mathbf{L} by a subject s having non-deductive reasons only to accept Γ and Λ .²⁷ s 's intermediate application of \mathbf{L} is such that s does not accept

²⁷Throughout this proof, in order to avoid excessive verbal clutter, I will sometimes let context disambiguate whether a sequence is accepted “conjunctively” (in the fashion of premises) or “disjunctively” (in the fashion of conclusions).

Δ, Ξ , even though she does accept, for every $\varphi \in \text{ran}(\Theta)$, Δ, φ . In such a situation, (NDD) only requires from s that she accept, for every $\varphi \in \text{ran}(\Theta)$, Δ, φ (since, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash_{\mathbf{L}} \Delta, \varphi$)—it does not require from s that she accept Δ, Ξ (since $\Lambda, \Gamma \not\vdash_{\mathbf{L}} \Delta, \Xi$). Therefore, s satisfies (NDD). However, in such a situation, (DD) does require from s that she accept Δ, Ξ . For, having non-deductive reasons to accept Γ , by (DD) s has deductive reasons to accept (and hence is committed to accepting), for every $\varphi \in \text{ran}(\Theta)$, Δ, φ (since, for every $\varphi \in \text{ran}(\Theta)$, $\Gamma \vdash_{\mathbf{L}} \Delta, \varphi$). By the additional principle of *semicolon-agglomeration of commitments to accepting “disjunctively”*:

(SACD) If, for some function seq from sentences to sequences, for every $\psi \in \text{ran}(\Pi)$ (with $\Pi = \chi_0, \chi_1, \chi_2 \dots$), one is committed to accepting “disjunctively” $\text{seq}(\psi), \psi$, then one is committed to accepting “disjunctively” $\text{seq}(\chi_0), \text{seq}(\chi_1), \text{seq}(\chi_2) \dots, \Pi$

(where ‘;’, unlike ‘,’ , denotes a right-conjunctive structural punctuation mark²⁸ and Π^\dagger the result of substituting \dagger throughout as the structural punctuation mark of Π), s is committed to accepting “disjunctively” $\Delta, \Delta, \Delta \dots, \Theta^\dagger$. Hence, by (W^r) , s is committed to accepting “disjunctively” Δ, Θ^\dagger .²⁹ Since s is also committed to accepting “conjunctively” Λ (having non-deductive reasons to accept “conjunctively” Λ), by the

²⁸A structural punctuation mark \dagger is *right-conjunctive* iff $[\Gamma \vdash \Delta \dagger \varphi \text{ iff } [\Gamma \vdash \Delta \text{ and } \Gamma \vdash \varphi]]$. Needless to say, the usefulness of a right-conjunctive structural punctuation mark derives from its ability to allow us to mimic conjunctive operations over sentences of a language which may well lack a conjunctive operator.

²⁹The use of (W^r) , a particular rule of *contraction*, is required in the proof only because, truth be told, our very official statements of left- and right-transitivity (T^l) and (T^r) are not as pure as one might wish but still encapsulate a mild form of contraction in their failure of letting $(\Gamma \text{ and } \Delta)$ vary depending on the value of ‘ φ ’. In view of the complexities that general precise statements of left- and right-transitivity free of any form of contraction would involve, I have however settled for sparing these to the reader as quite inessential to almost every point made in this essay and for only mentioning them on this occasion just to say that (W^r) would not be needed were they assumed in lieu of (T^l) and (T^r) .

additional principle of *monotonicity of commitment to accepting “disjunctively” over implication of commitment to accepting*:

(MCDIC) If commitment to accepting (“conjunctively”) Π_0, Π_1 implies commitment to accepting (“disjunctively”) Σ_0 , then commitment to accepting “conjunctively” Π_0 implies that commitment to accepting “disjunctively” Σ_1, Π_1 implies commitment to accepting “disjunctively” Σ_1, Σ_0 ,

we have that s is committed to accepting “disjunctively” Δ, Ξ (since the fact that $\Lambda, \Theta \vdash_{\mathbf{L}} \Delta, \Xi$ together with the reflexivity of \mathbf{L} , (NDD) and (DD) allows us to detach the main consequent of the relevant instance of (MCDIC)). Therefore, s does not satisfy (DD). Hence, (NDD) does not imply (DD), otherwise s could only satisfy the former by satisfying the latter.

- *Right-to-left.* By cases. Take a reflexive non-transitive consequence relation \mathbf{L} and suppose that, for every $\varphi \in \text{ran}(\Theta_0)$, a subject s has deductive reasons to accept φ . This can be so:

- (i) Either because s has non-deductive reasons to accept φ (so that, by reflexivity of \mathbf{L} and (NDD), s has deductive reasons to accept φ);
- (ii) Or because $\Gamma \vdash_{\mathbf{L}} \varphi$, where $\Gamma \neq \emptyset$ and s has non-deductive reasons to accept Γ (so that, by (NDD), s has deductive reasons to accept φ);
- (iii) Or because $\emptyset \vdash_{\mathbf{L}} \varphi$ (so that, by (NDD), s has deductive reasons to accept φ).

These are all the possible cases assuming, very plausibly, that it is not possible that, if $\emptyset \not\vdash_{\mathbf{L}} \varphi$, $[\Gamma$ is such that, for all (some) of its coordinates φ_0 , s has only deductive reasons to accept φ_0 , namely that φ_0 follows from premises Γ_0 , and that Γ_0 is in turn such that, for all

(some) of its coordinates φ_1 , s has only deductive reasons to accept φ_1 , namely that φ_1 follows from premises Γ_1 , and that Γ_1 is in turn such that, for all (some) of its coordinates φ_2 , s has only deductive reasons to accept φ_2 , namely that φ_2 follows from premises $\Gamma_2 \dots$]: in other words, the relation *x-but-not-y-is-a-coordinate-of-a-sequence-which-gives-s-deductive-reasons-for-y* must be well-founded on the field of logically contingent sentences. This yields that, if $\emptyset \not\vdash_{\mathbf{L}} \varphi$, each deductive reason s has to accept φ must ultimately be traceable back in a finite number of steps to an original sequence of premises Γ_i every coordinate of which s has non-deductive reasons to accept. Hence, given (T') , we have that $\Gamma_i \vdash_{\mathbf{L}} \varphi$. Letting $\Gamma = \Gamma_i$, this finite-chain case is finally reduced to case (ii).

Now, suppose also that $\Theta_0 \vdash_{\mathbf{L}} \Xi$. Let Θ_1 be the sequence obtained from Θ_0 by replacing each coordinate falling under case (ii) but not under case (iii) with one of its associated $\Gamma \neq \emptyset$ s has non-deductive reasons to accept and by deleting each coordinate falling under case (iii). Then, s has non-deductive reasons to accept all the coordinates of Θ_1 and, by (T') , $\Theta_1 \vdash_{\mathbf{L}} \Xi$ just as well. Hence, (NDD) gives s deductive reasons to accept Ξ .

□

Indeed, given that (DD) straightforwardly implies (NDD) no matter whether the consequence relation is transitive or not as long as it is reflexive (since non-deductive reasons to accept imply then deductive reasons to accept), theorem 5.4.1 can be strengthened to the effect that [(NDD) is equivalent with (DD) iff the consequence relation is transitive]. In view of this, it seems to me that the non-transitivist can reasonably insist that the pure principle, free of any assumption concerning formal properties of consequence, which represents its normativity (normativity which is the same for non-transitivists and transitivists alike) is (NDD) rather than (DD), (DD) being associated with such normativity only because equivalent with (NDD)

under the (rejected) assumption of transitivity.³⁰ Hence, even though a non-transitive logic is naturally hospitable to a certain “softening” of the normative force of consequence, this does not mean that no important requirement is placed by non-transitive consequence on rational beings. Indeed, as theorem 5.4.1 shows, the requirement placed by non-transitive consequence, (NDD), is just what becomes of the traditional requirement (DD) once the assumption of the transitivity of the consequence relation is dropped.

Before proceeding further, an absolutely essential feature of this dialectic must be made clear. (DD) is in effect a principle of closure of having deductive reasons to accept under logical consequence. Why then cannot one apply in this case a strategy analogous to the one we used in order to uphold in a non-transitive framework the necessity of truth preservation for consequence, which is in effect a principle of closure of truth under logical consequence? In the case of truth and of the other properties we considered in section 5.3.2, it was observed that there seem to be bridge principles linking the languages talking about these properties and the original language \mathcal{L}_0 claimed to be non-classical, principles which force the logic of the former languages to be

³⁰I should stress that, while it is clear that intermediate and strong applications of non-transitivity require rejection of (DD), it is much less clear that there is any interesting application of non-transitivity which does accept (DD). For a simple case study, consider a non-transitivist who assumes a sequence $\langle \mathbb{F}_0, \mathbb{F}_0, \mathbb{F}_1 \rangle$ to lack the left-transitivity property only if there can be no objectively well-grounded commitment to accept all the premises of an argument whose form is \mathbb{F}_0 (being e.g. motivated to do so by the desire of blocking **CONTRADICTION** *via* a non-transitivist strategy). (DD) (together with reflexivity of consequence and (NDD)) would still force on her the claim that, if one is committed to ‘ φ and it is not the case that φ ’, one is also committed to ψ , which seems to sit very uncomfortably with her claim that ψ does not follow from ‘ φ and it is not the case that φ ’. Note though that something akin to (DD) could actually be rescued from this perspective by replacing ‘deductive reasons’ with ‘objectively well-grounded deductive reasons’ in it: since, from this perspective, the only cases where a sequence $\langle \mathbb{F}_0, \mathbb{F}_0, \mathbb{F}_1 \rangle$ lacks the left-transitivity property are those where there can be no objectively well-grounded commitment to accept all the premises of an argument whose form is \mathbb{F}_0 , (DD) would be trivially satisfied in such cases (and satisfied in all other cases of consequence).

itself non-classical. No such principle seems to govern properties like one-has-deductive-reasons-to-accept- x and, more generally, there does not seem to be any reason why the language \mathcal{L}_1 talking about this property should exhibit the same problematic features which motivated a deviance from classical logic for \mathcal{L}_0 . Even for an intuitionist, for example, the language talking about which sentences of a standard quantified arithmetical language the members of a finite community of mathematicians have deductive reasons to accept may well be classical.

Moreover, even if the logic of \mathcal{L}_1 were non-transitive, (DD) would still be unacceptable for most non-transitivists. Suppose that a subject s has non-deductive reasons to accept φ , and that ψ follows from φ and χ from ψ . Suppose also that the sequence featuring as coordinates the forms of these two arguments (from φ to ψ and from ψ to χ) lacks the left-transitivity property, so that, *qua* non-transitivists, we would wish to avoid imputing to s any commitment to accepting χ , despite her having (non-deductive) reasons to accept φ . Now, by (NDD), we can infer that s has deductive reasons to accept (and hence is committed to accepting) ψ . But reflect that the satisfaction of s 's commitment to accepting ψ requires s to *treat ψ in a certain way* (at the least, very roughly, to assent to ψ if queried under normal circumstances). Crucially, such a way is also sufficient to establish s 's commitment to ψ independently from s 's having non-deductive reasons to accept φ together with (NDD). This means that there will typically be not only deductive reasons to believe that s is committed to accepting ψ (namely, those provided by s 's having non-deductive reasons to accept φ together with (NDD)), but also non-deductive reasons to believe so (namely, those provided by the observation of s 's behaviour). More precisely, there will be such reasons whenever s in fact satisfies her commitment to accepting ψ . Having non-deductive reasons both to accept that s is committed to accepting (and hence, possibly by reflexivity of consequence and (NDD), has deductive reasons to accept) ψ and to accept (DD), we would thus be committed to believing that s has deductive reasons to accept (and hence is committed to

accepting) χ ! (Of course, the argument propagates forward to any conclusion connected with χ through a chain of valid arguments.)

Actually, a very general lesson can be extracted from the main turn of the previous argument. Say that a proposition is *non-deductively inaccessible* for a subject s at time t iff its truth does not imply the availability for s at t of any non-deductive reason to believe it. What the previous argument shows is that even an intermediate application of non-transitivity with respect to two arguments α_0 and α_1 made by a subject s at time t requires the propositions expressed by the conclusions of α_0 to be non-deductively inaccessible for s at t . This squares nicely with the joint acceptance of (NDD) and rejection of (DD) typical of intermediate applications of non-transitivity: for what these imply is that whether or not the normative force of consequence applies to a subject's acceptance of a sentence should depend on the sentence's *pedigree* (non-deductive or deductive) in the subject's epistemic history, whereas lack of non-deductive inaccessibility precisely obliterates any distinction which might be drawn at that level.

5.4.4 The Asymmetry between Premises and Conclusions

Exploiting the distinction just unearthed, the left-hand and right-hand sides of a sequent can then be interpreted in a non-transitive framework as expressing a connection between what we have non-deductive reasons to accept and what we have deductive reasons to accept (which does not of course rule out that, independently, we also have non-deductive reasons to accept it: reasons-based acceptance can be overdetermined). This distinction has substance for transitivists and non-transitivists alike: there is a perfectly good sense in which I have non-deductive reasons to accept that snow is white but have simply deductive reasons to accept that either snow is white or grass is blue. It should thus actually be common ground that consequence can indeed fail to preserve non-deductive reasons to accept, and that it only

guarantees that, if one has non-deductive reasons to accept all the premises, one has deductive reasons to accept the conclusions.

Within this common ground, the debate on transitivity can then be understood as follows. The non-transitivist can be seen as thinking that consequence can fail to preserve deductive reasons to accept. She can thus be seen as focussing on the common-ground property of consequence of *guaranteeing* deductive reasons to accept *given* non-deductive reasons to accept, whilst the transitivist can be seen as focussing on the further (alleged) property of consequence of simply *preserving* deductive reasons to accept. The non-transitivist thinks that consequence guarantees a connection between the two different acceptance-related normative properties of being N_0 and of being N_1 , whilst the transitivist thinks that it also preserves the single acceptance-related normative property of being N_1 .

As a result of this, the non-transitivist's position can be summarized by the slogan that it takes more than being on the right-hand side of a valid sequent whose left-hand side coordinates are all accepted to figure as (one of the coordinates of) the left-hand side of a valid sequent whose other left-hand side coordinates are all accepted in such a way as to commit one to the latter sequent's conclusions. Sentences accepted *qua* conclusions of valid arguments all of whose premises are accepted may just not have the right pedigree to enter in turn as premises into further valid arguments possessing normative force: this peculiar relevance of the premises' pedigree to the normative force of a valid argument is where deductive non-transitive reasoning comes closest to inductive probabilistic reasoning. I suggest that this epistemic asymmetry between premises and conclusions standing in the consequence relation is the core of the non-transitivist's position.

It is worth noting that, parallel to the *epistemic* distinction between having non-deductive reasons to accept and having deductive reasons to accept, a similar distinction can be drawn at the *metaphysical* level between truth simply in virtue of how the world is and truth in virtue of how the world is and the laws of logic. The left-hand and right-hand sides of a sequent

can then equally legitimately be interpreted in a non-transitive framework as expressing a connection between what is true simply in virtue of how the world is and what is true in virtue of how the world is and the laws of logic (which does not of course rule out that, independently, it is also true simply in virtue of how the world is: truth can be overdetermined). This distinction too has substance for *dualists* and *non-dualists* alike: there is a perfectly good sense in which it is true simply in virtue of how the world is that snow is white, but it is true in virtue of how the world is and the laws of logic that either snow is white or grass is blue. It should thus actually be common ground that consequence can indeed fail to preserve truth simply in virtue of how the world is, and that it only guarantees that, if all the premises are true simply in virtue of how the world is, some conclusion is true in virtue of how the world is and the laws of logic.

The dualist can then be seen as thinking that consequence can also fail to preserve truth in virtue of how the world is and the laws of logic. She can thus be seen as focussing on the common-ground property of consequence of *guaranteeing* truth in virtue of how the world is and the laws of logic *given* truth simply in virtue of how the world is, whilst the non-dualist (either transitive or non-transitive) can be seen as focussing on the further (alleged) property of consequence of simply *preserving* truth in virtue of how the world is and the laws of logic. The dualist thinks that consequence guarantees a connection between two different metaphysical properties, whilst the non-dualist thinks that it also preserves a single metaphysical property.

5.4.5 The Locality of Non-Transitive Consequence

In the peculiar joint acceptance of (NDD) and rejection of (DD) we see in which sense the requirement placed by a non-transitive logic might typically be “*local*”, extending only so far as the consequences of sentences accepted for non-deductive reasons go rather than stretching to cover every consequence of any sentences one is committed to accepting for whichever reasons. Indeed, to

pursue this suggestive spatial metaphor³¹ even beyond the normative aspects of consequence, the non-transitivist's picture of the logical space is the rather unusual one which sees the space of consequences of a given point (that is, of a given collection of sentences) as being bounded by a *horizon*: there are bounds to what is logically necessary in relation to a given point which are not generally such in relation to themselves and to some other points included in the same original horizon, so that a movement inside the horizon can result in a movement of the horizon itself. Needless to say, this picture goes precisely against the more traditional picture of the logical space (inspiring the view of consequence as a *closure* operation introduced in fn 16) which sees the space of consequences of a given point as being bounded by a *frame*: there are bounds to what is logically necessary in relation to a given point which stretch so far as to be such also in relation to themselves and to all other points included in the same original frame, so that no movement inside the frame can result in a movement of the frame itself.

The picture that has emerged is then as follows. From the non-transitivist's perspective, at least as espoused in intermediate and strong applications of non-transitivity, the effects of consequence are peculiarly local: consequence manages to constrain the truth value of sentences only, as it were, *at one remove*. Again, to stress, this need not be because consequence in turn fails to impose a truth-preservation constraint on the sentences whose truth values have been so constrained (unless one is a dualist)—rather, it is because the application of such a constraint somehow fails to generate analogous effects. Rather than being manifested in the rejection of a law (i.e. a *proposition*) of truth preservation under logical consequence (unless one is a dualist), endorsement of the idea that consequence fails to constrain further truth values is thus manifested in the non-transitivist's pattern of attitudes of acceptance and non-acceptance (i.e. in her conforming or not to certain *rules* of acceptance and non-acceptance).³²

³¹Suggested to me in conversation by Crispin Wright.

³²Compare with the idea that consequence manages to guarantee the connection be-

The general condition which is thought by non-transitivists to trigger such a situation of merely local effects of consequence can very abstractly be thought of as one where transitivity would create *spurious connections*. To go back to the examples of applications of non-transitivity of section 5.2, in the case of relevance, transitivity would create a spurious *meaning* connection between sentences which have none (even though they are the opposite extremes of a chain of genuinely connected sentences); in the case of tolerance, transitivity would create a spurious *indiscriminability* connection between sentences which have none (even though they are the opposite extremes of a chain of genuinely connected sentences); in the case of probabilistic reasoning, transitivity would create a spurious *evidential* connection between sentences which have none (even though they are the opposite extremes of a chain of genuinely connected sentences). In all these cases, transitivity would obliterate a non-trivial distance structure which non-transitivists think is induced by consequence (possibly together with a theory) on a set of sentences, inflating local connections between these into global ones.

5.4.6 Non-Transitivist Theories, Situations and Worlds

It is in view of this asymmetry that, in a non-transitive framework and under a certain assumption to be introduced shortly, two different readings of theoretical notions defined using the notion of *closure under logical consequence* must be sharply distinguished as being non-equivalent (in this section, for simplicity's sake, we assume monotonicity and return to taking sets to be the terms of the consequence relation and to restricting our attention to single-conclusion arguments). Notions so defined can be seen as having at their

tween being the case and being *determinately* the case. Rather than being manifested in the acceptance of a law stating that, for every P , if P , then determinately P (which would spell disaster in its contrapositive form), that idea is manifested in one's pattern of attitudes of acceptance and non-acceptance, e.g. in her conforming to the rule of [accepting 'Determinately, φ ' whenever one accepts φ].

core the notion of a *theory*, defined as being any set of sentences closed under logical consequence. The rationale for the theoretical interest of a notion which, in virtue of the closure clause, outruns that of a mere set of sentences should be evident in view of the normativity of consequence. One has non-deductive reasons to accept a set of sentences X . By (NDD), one has thereby deductive reasons to accept (and hence is committed to accepting) not only X , but also the set of all the consequences of X . The objects of commitment are thus always closed under logical consequence in the specified sense. But theories are traditionally thought precisely to be the objects of commitment: theories are what people are traditionally thought to hold, defend, attack, revise, try to confirm etc.

Under the simplifying assumptions made at the start of this section, we can identify positions (introduced in section 5.4.3) with arbitrary sets of sentences; the theory of a position is then the closure under logical consequence of that position (let us denote by ‘ $\text{thr}_{\mathbf{L}}$ ’ the function from positions to theories under the consequence relation \mathbf{L}). A theory \mathcal{T} is *prime* iff, whenever ‘ φ or ψ ’ belongs to \mathcal{T} , either φ or ψ belongs to \mathcal{T} : primeness is the property of a theory to provide a witness for each of its assertions. A theory \mathcal{T} is *maximal* iff, for every φ , either φ or ‘It is not the case that φ ’ belongs to \mathcal{T} : maximality is the property of a theory to settle every question. A prime theory represents a *situation*, a maximal theory a *world*; φ is *true** in a situation (*true** in a world) iff φ belongs to the theory representing that situation (world). Call the logic under which a theory is closed ‘*target logic*’, the logic of the language talking about theories ‘*background logic*’.

What do non-transitivist theories, situations and worlds look like? There is no reason to think that the language talking about the logical consequences of sets of sentences of a language \mathcal{L} should exhibit the same problematic features which motivate a deviance from classical logic for \mathcal{L} . It behoves us then to consider the case where the background logic is classical. Under this assumption, two readings of the phrase ‘The theory of a position is the closure under logical consequence of that position’ must be sharply distinguished as

being non-equivalent:

- (i) The theory of a position P is the set of the logical consequences of P .
- (ii) The theory of a position P is the smallest set \mathcal{T} such that:
 - (a) \mathcal{T} contains all the logical consequences of P ;
 - (b) If φ is a consequence of \mathcal{T} , $\varphi \in \mathcal{T}$.³³

Under the assumption of transitivity of the background logic, the notion delivered by reading (ii) is clearly too strong even for intermediate applications of non-transitivity, since it will force the theory of a position—that to which one is committed—to contain sets of sentences which are not logical consequences of the position (this inadequacy of reading (ii) should come as no surprise given that it in effect amounts to an obliteration of the asymmetry between premises and conclusions which we have seen in section 5.4.4 to be crucial to intermediate and strong applications of non-transitivity).

Again, as in the case of (NDD) and (DD), it seems to me that, in her use of reading (i) rather than reading (ii), the non-transitivist is occupying a reasonable position, given that the following theorem holds for every reflexive consequence relation (see fn 34):

Theorem 5.4.2. *Reading (i) implies reading (ii) iff the consequence relation is transitive.*

Proof.

- *Left-to-right.* We prove the contrapositive and focus without loss of generality on failure of left transitivity. Take a reflexive non-transitive consequence relation \mathbf{L} such that, for every $\varphi \in Y$, $X \vdash_{\mathbf{L}} \varphi$ and $Z \cup Y \vdash_{\mathbf{L}} \psi$, but $Z \cup X \not\vdash_{\mathbf{L}} \psi$, and consider a position $P = Z \cup X$. Reading

³³That is, the glb under \subseteq of the class of sets satisfying (a) and (b); that is, the set \mathcal{T}_0 such that $\varphi \in \mathcal{T}_0$ iff, for every \mathcal{T}_1 satisfying (a) and (b), $\varphi \in \mathcal{T}_1$.

(i) dictates that $\psi \notin \text{thr}_{\mathbf{L}}(P)$ (since $Z \cup X \not\vdash_{\mathbf{L}} \psi$), whereas reading (ii) dictates that $\psi \in \text{thr}_{\mathbf{L}}(P)$ (since it dictates that $Y \subseteq \text{thr}_{\mathbf{L}}(P)$, and $Z \cup Y \vdash_{\mathbf{L}} \psi$).³⁴

- *Right-to-left.* Take a transitive consequence relation \mathbf{L} . Let $\text{thr}(\text{i})_{\mathbf{L}}$ and $\text{thr}(\text{ii})_{\mathbf{L}}$ be the functions corresponding respectively to readings (i) and (ii), and consider an arbitrary position P . Suppose that $\varphi \in \text{thr}(\text{ii})_{\mathbf{L}}(P)$. We aim to prove that $P \vdash_{\mathbf{L}} \varphi$. We do so by first defining by transfinite recursion the following hierarchy of positions:

- (i') $P_0 = \{\varphi : P \vdash_L \varphi\};$
- (ii') $P_{\alpha+1} = P_{\alpha} \cup \{\varphi : P_{\alpha} \vdash_{\mathbf{L}} \varphi\};$
- (iii') $P_{\lambda} = \bigcup(\{P_{\alpha} : \alpha < \lambda\}).$

Note that, since the language is finitary, by the well-ordering of the ordinals there will be a (not very big) first ordinal κ at which the process stabilizes and no new sentences are admitted as consequences—that is, for every $\alpha > \kappa$, $P_{\alpha} = P_{\kappa}$. Consider then the set $P_{\kappa+1}$. $P_{\kappa+1}$ satisfies (a), since $P_0 \subseteq P_{\kappa+1}$. $P_{\kappa+1}$ also satisfies (b), since $P_{\kappa+2} = P_{\kappa+1}$. Moreover, $P_{\kappa+1}$ is the smallest set to do so. For suppose that $\varphi \in P_{\kappa+1}$ and \mathcal{T} satisfies (a) and (b). We prove by transfinite induction that, for every $\alpha \leq \kappa + 1$, $P_{\alpha} \subseteq \mathcal{T}$:

- (i'') Since, by (a), $\{\varphi : P \vdash_L \varphi\} \subseteq \mathcal{T}$, $P_0 \subseteq \mathcal{T}$;
- (ii'') If $P_{\alpha} \subseteq \mathcal{T}$, since, by (b), $\{\varphi : P_{\alpha} \vdash_{\mathbf{L}} \varphi\} \subseteq \mathcal{T}$ as well, $P_{\alpha+1} \subseteq \mathcal{T}$;
- (iii'') If, for every $\alpha < \lambda$, $P_{\alpha} \subseteq \mathcal{T}$, then $\bigcup(\{P_{\alpha} : \alpha < \lambda\}) \subseteq \mathcal{T}$ as well.

Thus, since $\varphi \in P_{\kappa+1}$, $\varphi \in \mathcal{T}$. Being $P_{\kappa+1}$ the smallest set to satisfy (a) and (b), it follows that $P_{\kappa+1} = \text{thr}(\text{ii})_{\mathbf{L}}(P)$. We can now prove by transfinite induction that, given the transitivity of \mathbf{L} , if $\varphi \in \text{thr}(\text{ii})_{\mathbf{L}}(P)$

³⁴The assumption of reflexivity of \mathbf{L} is needed in order to ensure that $Z \subseteq \text{thr}_{\mathbf{L}}(P)$ as well as $Z \subseteq P$.

$(\varphi \in P_{\kappa+1})$, then $P \vdash_{\mathbf{L}} \varphi$. We do so by proving by transfinite induction the more general result that, for every $\alpha \leq \kappa + 1$, if $\varphi \in P_\alpha$, then $P \vdash_{\mathbf{L}} \varphi$:

(i''') If $\varphi \in P_0$, then $P \vdash_L \varphi$;

(ii''') If $\varphi \in P_{\alpha+1}$, then $\varphi \in P_\alpha \cup \{\varphi : P_\alpha \vdash_{\mathbf{L}} \varphi\}$, and so:

(a') Either $\varphi \in P_\alpha$, in which case, by the induction hypothesis, $P \vdash_{\mathbf{L}} \varphi$;

(b') Or $\varphi \in \{\varphi : P_\alpha \vdash_{\mathbf{L}} \varphi\}$, in which case, since, by the induction hypothesis, for every $\psi \in P_\alpha$, $P \vdash_{\mathbf{L}} \psi$, by (T') $P \vdash_{\mathbf{L}} \varphi$ as well;

(iii''') If $\varphi \in P_\lambda$, then, for some $\alpha < \lambda$, $\varphi \in P_\alpha$, and so, by the induction hypothesis, $P \vdash_{\mathbf{L}} \varphi$.

The proof is completed by observing that, since $P \vdash_{\mathbf{L}} \varphi$, $\varphi \in \text{thr}(\mathbf{i})_{\mathbf{L}}(P)$ as well.

□

Indeed, given that reading (ii) straightforwardly implies reading (i) no matter whether the consequence relation is transitive or not (since, given (a), it is trivial that $[\varphi \in \text{thr}(\mathbf{i})_{\mathbf{L}}(P) \text{ only if } \varphi \in \text{thr}(\mathbf{ii})_{\mathbf{L}}(P)]$), theorem 5.4.2 can be strengthened to the effect that [reading (i) is equivalent with reading (ii) iff the consequence relation is transitive]. In view of this, it seems to me that the non-transitivist can reasonably insist that the pure notion, free of any assumption concerning formal properties of consequence, which plays the complex role usually assigned to the notion of a theory (role which is the same for non-transitivists and transitivists alike) is the one expressed by reading (i) rather than the one expressed by reading (ii), the latter being associated with such role only because reading (ii) is equivalent with (i) under the (rejected) assumption of transitivity

Once reading (i) is distinguished as the appropriate notion to use when the target logic is non-transitive, the theory of theories can proceed very

much as before (see e. g. Barwise and Perry [1983], pp. 49–116). For our purposes, we only need to note the following concerning situations. A prime theory \mathcal{N} with a non-transitive target logic \mathbf{L} representing a certain situation may be such that $\psi \in \mathcal{N}$, $\psi \rightarrow \chi \in \mathcal{N}$ but $\chi \notin \mathcal{N}$ (such will be the case consider for example in tolerant logics). The situation thus represented would be one where a conditional and its antecedent are true*, but the consequence is not. At a glance, this may of course look like an unwelcome consequence, since \mathbf{L} may actually be such that the rule of *modus ponens* holds in full generality (as it does in tolerant logics). Even worse, if the law of excluded middle also holds unrestrictedly in \mathbf{L} (as it does in classical tolerant logics), given \mathcal{N} 's primeness we would have that 'It is not the case that χ ' $\in \mathcal{N}$! However, it should by now be clear that, analogously to the cases discussed in section 5.3.2, these consequences are due to the choice of adopting a transitive background logic in the theory of theories. This choice imposes that the link between the technical notion of truth* in a situation (and of truth* in a world) and the informal and philosophical notion of truth in a situation (and of truth in a world) be at best very complex and mediated.

Again, the point can be illustrated with reference to more well-known deviations from classical logic. Consider a classical theory \mathcal{C} of a prime intuitionist theory \mathcal{I} formulated in language \mathcal{J} . \mathcal{C} entails that the situation i represented by \mathcal{I} is actually such that, for every $\varphi \in \mathcal{J}$, either φ is true* in i or φ is not true* in i (even though, of course, being the target logic intuitionist, it need not be the case that, for every $\varphi \in \mathcal{J}$, either $\varphi \in \mathcal{I}$ or 'It is not the case that φ ' $\in \mathcal{I}$). Such a conclusion would be repugnant given what would seem to be the most natural theory of truth in a situation available to an intuitionist (namely one such that φ 's being not true in a situation implies that φ is false in that situation and so that 'It is not the case that φ ' is true in that situation).

If a non-transitive logic is adopted as background logic, however, things change drastically and reading (ii) becomes again a viable option for intermediate and strong applications of non-transitivity. Indeed, the following

theorem concerning preservation of truth^* in a situation becomes available, at least for those applications of non-transitivity which (like the strong application of tolerant logics to the sorites paradox) do not require failures of transitivity for standard definitional reasoning:

Theorem 5.4.3. *If the background logic \mathbf{L}_0 is non-transitive, reading (ii) allows for the construction of a theory \mathcal{T} with a non-transitive target logic \mathbf{L}_1 representing a situation s such that, if $X \vdash_{\mathbf{L}_1} \varphi$, then truth^* in s is preserved from X to φ .*

Proof. Given that we are only concerned with applications of non-transitivity which do not require failures of transitivity for standard definitional reasoning, it suffices to observe that the following argument only requires transitivity in that area. Suppose that $X \vdash_{\mathbf{L}_1} \varphi$ and that every member of X is true^* in s . Then, by definition of truth^* in s , it follows that $X \subseteq \mathcal{T}$. From this, $X \vdash_{\mathbf{L}_1} \varphi$ and reading (ii), it follows that $\varphi \in \mathcal{T}$, which in turn entails, by definition of truth^* in s , that φ is true in s as well.

□

On the other hand, going back to the theory \mathcal{N} discussed two paragraphs back, reading (ii) no longer has the unwelcome consequence of declaring that $\chi \in \mathcal{N}$. For, in a non-transitive background logic, that $\chi \in \mathcal{N}$ does not follow from set theory, reading (ii), φ 's belonging to \mathcal{N} , $\varphi \rightarrow \psi$'s belonging to \mathcal{N} and $\psi \rightarrow \chi$'s belonging to \mathcal{N} .

5.5 Conclusion

Appealing as it is, the non-transitivist solution to the sorites paradox requires an extended philosophical discussion of what sense there is to be made of a rational being who reasons using a non-transitive logic. This chapter has no doubt been only a first stab at meeting that pressing request in its generality. We have seen how this task involves dealing with some of the hardest

problems at the interface between the philosophy of logic and the theory of rationality: the connection between structural rules and the very nature of the premises and conclusions of an argument (and of their acceptance or rejection); the relation between deductive and non-deductive consequence relations; the relation between consequence and inference; the relation between consequence, preservation of truth and preservation of other epistemic properties; the normativity of consequence; the relation between premises and conclusion of a valid argument; the nature of the object of logical commitment; the problem of theorizing about how all this behaves in a certain logic by using a different one etc. From a wider perspective, I hope that the foregoing investigations about how non-transitivity impinges on these and other issues have helped to see them in a new, more general light, and that the conceptualizations made and the distinctions drawn will prove fruitful also in the examination of the philosophical foundations of other logics. As far as the more specific purposes of this essay are concerned, I take these investigations, applying as they do in particular to tolerant consequence relations, to provide a solid philosophical basis to the non-transitivist solution to the sorites paradox, solution which is in my view the best one available to the naive theory of vagueness defended in this essay. Within the restricted focus of this essay, it remains to be seen whether and how this machinery can be brought to bear on a non-inferential counterpart of the sorites paradox. In the next chapter, I will offer a mainly negative answer to this question, answer which will ultimately allow us to explain nicely within the naive theory the phenomenon of vagueness which has so far stayed impenetrable to it—borderlineness.

Chapter 6

Forced March in the Penumbra

6.1 Introduction and Overview

Oftentimes, a ‘Yes’/‘No’-answer to a ‘Yes’/‘No’-question is not warrantedly available from one’s current epistemic state. Both by replying ‘Yes’ and by replying ‘No’ one would be returning a verdict which would not be warranted by the lights of one’s overall epistemic situation. Such would be a ‘Yes’/‘No’-answer, I take it, to the question as to whether the number of electrons in the universe is odd. If one were to reply ‘Yes’ or ‘No’ to such a question, one would be judging¹ something for which no epistemic (ie, very broadly, truth conducive) support is available—since, epistemically, a serious judgement (as opposed to a conjecture, or supposition, or phantasy etc.) does require some positive epistemic support or other, one would thus be doing something *unwarranted* and be consequently *criticizable*.

To be sure, the criticism would still be (at least immediately) purely *intel-*

¹Throughout, I will use ‘judge’ and its like more or less interchangeably with ‘believe’ and its like, with a preference for the former when emphasizing the “*act*” aspect of the relevant mental event (for example, when talking about the mental events in some sense expressed by speech acts such as answers) and a preference for the latter when emphasizing the “*state*” aspect of the relevant mental event.

lectual—it would only (at least immediately) concern how well one is doing in one's role as *inquirer* (searcher of truth), not how well one is doing in one's overall *personal* life. Even with respect to one's intellectual life, the criticism would only be straightforwardly licensed by the consideration of the judgement as candidate for being part of the *final* good one should strive for in one's intellectual life—the judgement could still constitute a precious *instrumental* good to that very same end (consider a situation where an answer is elicited on pains of a sudden and certain death). An overall intellectual criticism, let alone an all-things-considered criticism, would thus surely need to take into account further aspects of the inquirer's total situation before being issued properly. Yet, the restricted intellectual criticism would remain in any case, and, other things being equal, could not but extend to an overall intellectual and indeed all-things-considered criticism. In the following, the use of normative and evaluative vocabulary is to be understood in the restricted intellectual sense.

Fortunately, 'Yes' and 'No' are not the only two answers available to a 'Yes'/'No'-question. The simple case just considered indicates the availability of a fall-back answer, which does not take a stand on the first-order question as to whether *P*, but simply reflects the inquirer's inability to return a warranted first-order verdict either way: 'I can't say', interpreted as implying that the inquirer neither is warranted in judging that *P* nor is warranted in judging that it is not the case that *P*. In many cases, such a fall-back seems to be readily available, allowing the inquirer to preserve her *intellectual integrity*: 'Did Socrates sneeze on his 28th birthday?', 'Is the continuum hypothesis true?', 'Is abortion permissible?'—'I can't say'. It is then natural to conjecture that the fall-back is always so available, and thus that life is relatively easy even for a *non-omniscient* subject (a searcher of truth, an inquirer). Granted, her view may be widely covered by the veil of ignorance—still, even in these unfortunate circumstances, if she is reflective enough, she can live up to any challenge posed by a 'Yes'/'No'-question by returning a warranted verdict relevant to the dialectical situation in which

the question is asked: ‘I can’t say’.² Not even the most cunning dialectical opponent could rob her of her intellectual integrity. Or could he?

The rest of the chapter is organized as follows. Section 2 undertakes a couple of qualifications and refinements necessary before our leading question can be properly examined. Drawing on results from two quite different areas, section 3 shows how no simple-minded version of the preservation of one’s intellectual integrity can be upheld and develops an elaborate version thereof. Taking inspiration from a case widely discussed in the contemporary debate on vagueness, section 4 introduces a major difficulty for the elaborate version and argues that no happy-face solution to it succeeds. Section 5 proposes an unhappy-face solution to the difficulty, tests its fruitfulness by applying it to another area and uses the materials introduced by it to reconstruct, within the traditional approach and more specifically within the rejection of the existence of strong borderline cases, a notion of a borderline case which accounts for the phenomenon of borderlineness. Section 6 draws the conclusions which follow from the overall dialectic for the connection between tolerance and borderlineness from the standpoint of the theory of vagueness defended in this essay.

²Something along these lines (with focus on knowledge) is arguably presupposed for example in the medieval practice of *obligatio* (see Dutilh-Novaes [2007], pp. 145–214), where, roughly, starting from an initial proposition accepted for the sake of argument, one is confronted sequentially with a series of propositions, each of which one has either to accept or to reject or to doubt (if one has not to distinguish senses because of the presence of an ambiguity) while preserving consistency. If a proposition is logically independent from one’s previous commitments, one is supposed to [accept it iff one knows it to be true] and [reject it iff one knows it to be false]. In such a case, one is presumably supposed to be knowledgeable about one’s ignorance if one decides to doubt the proposition. If so, it is in effect presupposed that, for every proposition *P*, one either knows that *P* or knows that it is not the case that *P* or knows that one does not know whether *P*. Thanks to Stephen Read for introducing me to the intriguing world of *obligationes*.

6.2 Preliminaries

6.2.1 A Simple-Minded Conjecture and Innocent Questions

Of course, ‘warrant’ and ‘I can’t say’ are not much more than place-holders in the foregoing considerations. Letting ‘*E*’ stand schematically for any significant epistemic property, the general form of the previous *unqualified and simple-minded* version of the conjecture is:

(US) For every ‘Yes’/‘No’-question *Q*, if neither ‘Yes’ nor ‘No’ would be an *E* answer to *Q* for subject *s* at time *t*, then ‘Neither ‘Yes’ nor ‘No’ would be an *E* answer to *Q* for me now’ would be an *E* answer to *Q* for *s* at *t*.

Let me stress from the very beginning that conditions like (US) (and its successors to come) will always be meant only as a first *approximate* attempt at stating a non-trivial sufficient condition for the preservation of one’s intellectual integrity in the face of dialectical or theoretical challenges.³ The concept of intellectual integrity is (unsurprisingly!) very complex, and the simple letter of the conditions to be considered may well fall short of capturing a real sufficient condition, allowing for deviant satisfaction. This is why in the following I will often take care of enriching the condition under discussion according to the specific demands which seem to be placed by the concept of intellectual integrity in a particular dialectical or theoretical context. It will also be one of the main points of my positive proposal that

³I am using ‘dialectical’ to refer to the task of *defending a view against other views*, ‘theoretical’ to refer to the task of *determining which view is true* (see Pryor [2004] for a recent interesting case-study investigation of this distinction). In this sense, there plausibly are dialectical norms which are not theoretical norms and *vice versa*. As indicated in the text, I will be interested in the issue of the preservation of one’s intellectual integrity in both dialectical and theoretical contexts, offering arguments which, given their insensitivity to this distinction, concern both.

nothing like these conditions is a *necessary* condition for the preservation of one's intellectual integrity in the face of dialectical or theoretical challenges.

Another crucial qualification, already implicit in the foregoing, must be entered right at the outset. This concerns epistemic properties—such as knowledge, reliability, externalist conceptions of warrant and justification etc.—whose exemplification plausibly depends on conditions that are not always subjectively detectable *at least in their failure to hold*. For any such property of being *E*, of course, there is no hope for (US) to be true, as one can easily imagine a situation where a subjectively undetectable condition fails to hold and not only do the inquirer's answers not exemplify *E*ness with respect to the relevant proposition that *P* and its negation, but the inquirer also mistakenly believes that *P* (mistakenly believes that it is not the case that *P*), and so returns a 'Yes' ('No') verdict which is not *E*. Given the potential negative subjective undetectability of some necessary condition for *E*ness's exemplification, one might simply be tricked into such a situation. *E*ness might for example be the property of being reliable with respect to the proposition that the Tower of Pisa is straight, and one might be fed (by perception, memory or testimony) false information—which seems by all appearances to come from a reliable source—to the effect that the Tower of Pisa is straight. One's 'Yes'-answer would then not be reliable.

However, such a possibility—even though sadly real—does little to endanger the ideal of intellectual integrity governing inquiry: false information need not always carry its name on its sleeves, and, at least in some cases, one might not be epistemically criticizable for taking it at face value. Indeed, as should already be clear, falsity is not necessary for triggering answers lacking *E*ness on the model of the above example: all that is needed is a piece of information (truthful or otherwise) which might be taken by the inquirer to support an *E* answer when in fact it does not (call any such piece of information '*faulty*'). *E*ness might for example be the property of knowing with respect to the proposition that the Tower of Pisa is crooked, and one might be fed (by perception, memory or testimony) faulty (though truthful)

information—which seems by all appearances to come from a knowledgeable source—to the effect that the Tower of Pisa is crooked. One’s ‘Yes’-answer would then not be knowledgeable.

We can conclude that the threat to an inquirer’s intellectual integrity seems only to come from questions targeting propositions *on which no faulty information bears which may be possessed by the inquirer* (call any such question ‘*innocent*’). Nothing as strong as (the false) (US) is needed to protect an inquirer’s intellectual integrity from innocent questions—the simple-minded conjecture can accordingly be *qualified* to:

(QS) For every innocent ‘Yes’/‘No’-question Q , if neither ‘Yes’ nor ‘No’ would be an E answer to Q for subject s at time t , then ‘Neither ‘Yes’ nor ‘No’ would be an E answer to Q for me now’ would be an E answer to Q for s at t .

We can also hereafter stick to the use of ‘warrant’ and its like to stand schematically for a wide range of significant epistemic properties, since, as will be seen, the arguments to be considered are quite insensitive in this respect. Indeed, they are at their strongest with broadly internalist properties, which are the most plausible candidates for spelling out the requirement of intellectual integrity which would be met were the corresponding instance of (QS) true. In all cases, we will focus for simplicity’s sake on the “propositional” rather than the “doxastic” version of the relevant epistemic property (see Firth [1978], p. 218 for the introduction of the distinction with respect to warrant properly called and for an influential explication of it).

6.2.2 Saying ‘No’

Before starting our critical discussion, another crucial clarification needs to be made. In a certain respect, the most natural setting for our problem—the *dialogical* setting of *questions* and *answers*—brings in unnecessary complexities. Presumably, answering ‘Yes’ to a question as to whether P in some

sense expresses a judgement to the effect that P . However, it is rather doubtful that an analogous situation holds for a ‘No’-answer—that is, it is rather doubtful that answering ‘No’ to a question as to whether P in some sense always expresses a *judgement* to the effect that it is *not* the case that P . A ‘No’-answer would rather seem to express a *rejection* of the proposition that P (see chapter 2, fn 7). The failure of the implication from a rejection to the corresponding negative judgement is exemplified for discourses and questions which are thought to reflect some kind of *defectiveness*, such as, on some theories of vagueness and truth (see e.g. Field [2003a]; Field [2003b]), questions asking whether a borderline red is red or whether the Liar sentence is true⁴—for a less controversial example of such defectiveness, think of a question asking whether it is the case that abracadabra. (Dialetheism, as defended in Priest [2006a]; Priest [2006b], offers a counterexample to the converse implication.)

Transposed into a reflective, non-dialogical setting, our problem would concern the warrantability of the attitudes a thinker may eventually adopt towards the proposition that P once she has become aware of the issue as to whether P and has reflected on it. In such a setting, a ‘No’-answer to the question as to whether P would have to be variously represented as either an attitude of acceptance that it is not the case that P or as an attitude of rejection of the proposition that P . In the following, I propose however to stick to the vividness of the dialogical setting and to impose the following conventions: a ‘No’-answer to a question as to whether P will be interpreted restrictedly (as expressing a negative judgement that it is not the case that P), while an ‘I can’t say’-answer, unless otherwise specified, will be used as an umbrella answer for expressing either a judgement to the effect that there is no warrant for judging whether P or a rejection of the proposition that P . An ‘I can’t say’-answer expressing a warranted rejection of a certain proposition

⁴These theories’ treatment of higher-order borderline cases and of the semantic paradoxes also belie the reduction of rejection that P to negative judgement that it is not true that P (that ‘ P ’ is not true).

will also be understood among the legitimate satisfiers of (QS) (and, in the following, of (E)) and, for simplicity's sake, we will take a rejection of the proposition that P to be warranted just in case the corresponding judgement that the proposition that P should be rejected is warranted.

6.3 Can't Say 'Can't Say'

6.3.1 Failure of Negative Transparency

Natural as it may be, (QS) comes immediately under pressure from at least two different areas. In both cases, I will first present the argument as it arises in relation to knowledge, since such a presentation is by far more familiar and natural. I will then point out the modifications needed, if any, to adapt the argument to the more general notion of warrant we are using.

The first area of pressure is constituted by the structural features illustrated by some of the *counterexamples against the characteristic S5 axiom for knowledge* (if one does not know that P , then one knows that one does not know that P). One's evidence might not suffice to know that P , and still be quite close to be so sufficient, so close that one does not have enough second-order evidence to know that the first-order evidence is so insufficient. In this case, just as a 'Yes' or a 'No'-answer, an 'I can't say'-answer to the question as to whether P would not be knowledgeable. The situation can be illustrated by the following example:

JACOB Jacob is about to conduct an experiment but, given the poor status of his knowledge of chemistry, is agnostic about its real bearing on the question as to whether there is calcium in a sample of water, whereas a positive result would actually come very close to settle the question in the positive, with only a further minor check due for knowledge to be delivered. Jacob is also agnostic as to whether he believes—on the basis of this or other experiments—that there is calcium in the

sample.⁵ If the experiment were then conducted yielding a positive result, Jacob would still not know that there is calcium in the sample, but he would also not know that he does not know, given his lack of evidence concerning the real bearing of the experiment.

Suppose then that, after the experiment has been carried out with a positive result, Jacob is asked whether there is calcium in the water sample. As **JACOB** has been described, neither a 'Yes'- nor a 'No'-answer would be knowledgeable, but neither would be in this case an 'I can't say'-answer (of course, it would not be so even if interpreted as expressing rejection of the proposition that there is calcium in the sample). An analogous point can be made by substituting throughout the notion of being warranted for that of knowing.⁶

6.3.2 Epistemic Paradox

Pressure against (QS) also comes from the area of the paradoxes of self-reference, in particular from the *epistemic paradoxes*. Consider a modified version of the *Knower paradox* (re-discovered in modern times by Kaplan and Montague [1960]), with a sentence κ provably equivalent with ' κ is not known by Thomas'⁷ (where 'know' and its like are short for 'know at some

⁵Assuming that (Jacob knows that) knowledge requires belief, this twist is needed in order to foreclose Jacob the easy route to knowledge of his ignorance which goes through knowledge of his lack of belief. The twist is not needed if the notion of being in a position to know is used instead of the notion of knowing, nor is it needed if the operative notion is that of being (propositionally) warranted.

⁶Many counterexamples against the characteristic **S5** axiom for an epistemic operator crucially rely on the possession of faulty information on the part of the subject (see e.g. Williamson [2000b], p. 23). **JACOB** shows that such reliance is dispensable.

⁷To keep things simple, in making this point I use 'know' and its like as predicates of sentences (and subjects). This is of course in contrast with much traditional philosophical thinking on this subject and, more generally, on the attitudes, which construes propositional-attitude verbs either as predicates of non-linguistic propositions (and sub-

time or other', with suitable modifications under embeddings), and suppose that Thomas is asked whether κ is known by Thomas.

- (i) Suppose that Thomas is knowledgeable in saying 'Yes'. Then he knows that κ is known by Thomas. By factivity of the outermost (sentential-operator) occurrence of 'know', κ is known by Thomas. Since by construction κ entails ' κ is not known by Thomas', it follows that, by factivity of knowledge (in the form 'If ' P ' is known by ξ , then P '), κ is not known by Thomas. Contradiction (by structural contraction).
- (ii) Suppose that Thomas is knowledgeable in saying 'No'. Then he knows that κ is not known by Thomas. Switching from the sentential-operator reading to the sentence-predicate reading, ' κ is not known by Thomas' is known by Thomas. Since by construction ' κ is not known by Thomas' entails κ , it follows that, by closure of knowledge under logical consequence (in the form 'If ' P_0 ', ' P_1 ', ' P_2 '... are all known by ξ , and they entail ' Q ', then ' Q ' is known by ξ '),⁸ κ is known by Thomas. Contradiction (by structural contraction).
- (iii) Suppose that Thomas is knowledgeable in saying 'I can't say'. Then he knows that he does not know that κ is not known by Thomas. Since by construction ' κ is not known by Thomas' is equivalent with

jects) or as term-indexed sentential operators ("connectives", as Prior [1971], p. 135 felicitously put it). I trust that this circumstance will not affect the substance of the point I'm about to make, since the very same point could have been made just as well, even though more clumsily, with a knowledge sentential operator together with the use of a truth predicate, or of propositional quantification, or of non-standard quotation and abbreviation. An analogous comment applies to the use of 'warrant' and its like below.

⁸Many authors reject this principle and accept only the weaker principle of closure of knowledge under *known* logical consequence. If only this weaker principle is accepted, the example must be enriched, here and elsewhere, with the further assumption that Thomas knows that the entailments in question holds. Thanks to Stephen Read for pointing out the need for this qualification, which will be left implicit for similar cases in the rest of this chapter.

κ , it follows that, by closure of knowledge under logical consequence (sentence-predicate form), Thomas's not knowing ' κ is not known by Thomas' is equivalent with Thomas's not knowing κ . Switching again from the sentential-operator reading to the sentence-predicate reading and using closure of knowledge under logical consequence (sentential-operator form), Thomas knows that κ is not known by Thomas. The argument to contradiction can then proceed as in (ii).

Again, neither a 'Yes'- nor a 'No'-answer would be knowledgeable, but neither would be in this case an 'I can't say'-answer. Moreover, since the argument in (iii) only exploits at the relevant places the formal property of closure under logical consequence of 'is known by Thomas', and since this property is plausibly exhibited also by 'should not be rejected by Thomas' (where 'reject' and its like are short for 'reject at some time or other', with suitable modifications under embeddings), the argument in (iii) can still conclude to Thomas's knowing that κ should be rejected by Thomas in case Thomas's answer 'I can't say' expresses rejection (rather than ignorance) of the proposition that κ is not known by Thomas. From there we cannot of course argue as before (connecting immediately to (ii)), but we appeal instead to the following constraint linking *rejection* and *ignorance*:

(RI) If φ should be rejected by a subject s , φ is not known by s

and argue as follows:

(iii') Suppose that Thomas is knowledgeable in saying 'I can't say'. Then he knows that ' κ is not known by Thomas' should be rejected by him. Since by construction ' κ is not known by Thomas' is equivalent with κ , it follows that, by closure of 'should not be rejected by Thomas' under logical consequence, ' κ is not known by Thomas's having to be rejected by Thomas is equivalent with κ 's having to be rejected by Thomas. By closure of knowledge under logical consequence

(sentential-operator form), Thomas knows that κ should be rejected by Thomas. By Thomas's knowledge of (RI) and closure of knowledge under logical consequence (sentential-operator form), Thomas knows that κ is not known by Thomas. The argument to contradiction can then proceed as in (ii).

An analogous point can be made by substituting throughout the notion of being warranted for Thomas for that of being known by Thomas and undertaking the appropriate adjustments (where 'warrant' and its like are short for 'warrant at some time or other', with suitable modifications under embeddings). In particular, in (i) we can no longer appeal to factivity. We thus assume the additional principles of *harmony of warrant*:

(HW) If φ is warranted for a subject s , ' φ ' is not warranted for s ' is not warranted for s

(in effect, a factivity principle for warrant restricted to ' φ ' is not warranted for s '-initial sentences) and of *unwarrantability of absurdity*:

(UA) If φ is absurd, φ is not warranted for any subject.

We also make use in all three cases of mild closure-under-logical-consequence principles for warrant. We consider the sentence F which is provably equivalent with ' F is not warranted for Thomas', and ask Thomas whether F is warranted for him. (i) now goes into:

(i'') Suppose that Thomas is warranted in saying 'Yes'. Then he is warranted in judging that F is warranted for Thomas. Since, by construction F entails ' F is not warranted for Thomas', ' F is warranted for Thomas' entails, by closure of warrant under logical consequence (sentence-predicate form), ' F is not warranted for Thomas' is warranted for Thomas'. By (HW), these are jointly absurd, and so, switching from the sentential-operator reading to the sentence-predicate

reading, it follows that, by closure of warrant under logical consequence (sentence-predicate form), an absurdity is warranted for Thomas. Contradiction with (UA).

(ii), (iii) and (iii') receive their natural modifications by substituting 'warrant' and its like for 'know' and its like. In the modification of (iii'), an analogue of (RI) for 'warrant' must be used.

While refuting (QS), these kinds of examples also show the way to a more acceptable version of a conjecture whose truth would guarantee the preservation of an inquirer's intellectual integrity. For reflect that, as far as they have been described, both kinds of examples are compatible with an 'I don't know that I know'-answer's being knowledgeable. Of course, this is not the end of the story as both kind of examples can be developed (or others be produced) where not even such an answer would be knowledgeable, but the spirit of the conjecture, allowing the inquirer to climb up as far as needed the hierarchy of the relevant epistemic property, should be clear enough.

A satisfactory precise formulation of the conjecture would certainly require considerable subtleties. For example, we don't want the requirement of intellectual integrity to be so low that it can be met, with respect to the question as to whether P , by the answer 'I don't know that [I know that P and I know that it is not the case that P]'. Such an answer would seem to change the subject (from whether P to logic) rather than to address the original question whether P and the challenge posed by it. Intellectual integrity is no less jeopardized by a willful change of subject than by an unwarranted answer—by so changing the subject, one would simply be cheating. A classification of states occurring in the hierarchy of warrant states into *admissible* and *inadmissible* ones would thus seem to be necessary for a satisfactory formulation of the conjecture. The task of stating such a classification appears to be highly non-trivial. Fortunately, the problem I wish to discuss in the rest of this chapter is wholly insensitive to this and other fine-tuning details, and we can rest content with the following (rather promissory) elaborate version

of the conjecture:

- (E) For every innocent ‘Yes’/‘No’-question Q , for some Φ that stands for an admissible state occurring in the hierarchy of warrant states, if neither ‘Yes’ nor ‘No’ would be a warranted answer to Q for subject s at time t , then ‘Neither ‘Yes’ nor ‘No’ would be a Φ answer to Q for me now’ would be a warranted answer to Q for s at t

(note the order of the quantifiers here, which allows the value of ‘ Φ ’ to be a function of the value of ‘ Q ’).

To see why (E) may be thought at least to stand a chance to deal adequately and exhaustively with the problems just highlighted for (QS), it is useful to think about the structurally similar situation arising within the dominant approach to vagueness. Within this approach, for some object between the definite cases of F ness and the definite cases of non- F ness, (first-order) vagueness determines that neither an ‘ F ’-predication nor a ‘non- F ’-predication should be accepted as *definitely* correct. The introduction of a new category, borderline F ness, allows one to *decide* (i.e. definitely categorize) some new cases which were not decidable by simply using the categories of F ness and non- F ness. Still, higher-order vagueness determines that not every case can be so decided. The introduction of a new category, borderline borderline F ness, allows one to decide some new cases which were not decidable by simply using the categories of F ness, non- F ness and borderline F ness. And so on, in such a way that one might reasonably conjecture that every case will become eventually decidable at some stage or other of the hierarchy, given the richer categorical discriminations afforded by each new stage. A structurally similar conjecture seems to be equally reasonable in our situation, and such a conjecture is just what (E) in effect amounts to.⁹ However, reasonable as (E) may seem, we now turn to a case which puts considerable pressure even on it.

⁹Thanks to Crispin Wright for intense questioning which led to this comparison.

6.4 The Forced March

6.4.1 Setting

Consider a series of 1,000 pairwise indiscriminable colour patches, going from clear red to clear orange. The patches are pairwise indiscriminable in the sense that, given any pair of adjacent patches, no normal human subject could detect, by unaided vision, any difference in colour between the two patches when considering them in isolation from all other patches (either sequentially or at the same time). Consider a normal human subject, Vicky, who, under the best visual conditions, is forced to give her best judgement sequentially on all the patches, starting from patch #1. Vicky may know about all this situation, and so know that at least some of the indiscriminable patches x and y could differ from one another at least in the sense of there being another patch z in the series such that z is discriminable from x but not from y (although the existence of such a patch is not guaranteed in the series). Still, Vicky cannot detect, by unaided vision, any difference in colour between any two adjacent patches when considering them in isolation from all other patches. The question that she will be asked at each step is ‘Is this patch red?’. Vicky’s aim is to give a warranted answer at each step of the forced march.¹⁰

To this effect, she has available a wide range of answers, not necessarily restricted to those allowed by (E): ‘Yes’, ‘No’, ‘I can’t say’, ‘It’s borderline’, ‘Yes and no’ etc.¹¹ However, for the same reason as (E) requires a distinction between admissible and inadmissible warrant states, we need to impose some

¹⁰The problem of vagueness actually entered in Western philosophy in something like this forced-march setting (see the references in chapter 1, fn 9 for a presentation and discussion of the sources). As far as I know, the forced march *with its specific problems* has been introduced in the contemporary debate on vagueness by Sainsbury [1992] (see Williamson [1994], pp. 8–27; Raffman [1994]; Raffman [1996]; Horgan [1994]; Tye [1994], pp. 204–6; Soames [1999], pp. 203–27; Fara [2000]; Priest [2003]; Shapiro [2003]; Shapiro [2006], pp. 1–44; Wright [2007c] for a representative range of approaches to the issue).

¹¹Note that it is no real limitation to our inquiry into the possibility and limits of intel-

minimal constraints on the answers available to Vicky, otherwise she would be able to go through the forced march simply by answering e.g. ‘I’m hungry’ at each step. Clearly, such a willful change of subject would be just as bad as an unwarranted answer as far as Vicky’s intellectual integrity is concerned—by so changing the subject, Vicky would simply be cheating. To state the constraints, we need a notion of an ordering of the propositions available to Vicky as answers according to what intuitively is their *informational strength*. Such an ordering will properly include the ordering of *logical strength* and will be such that propositions entailing that the patch in question is red (not red) are stronger than any other proposition save for those entailing that the patch in question is not red (red). Fortunately, we can remain rather vague on the rest, leaving it at the intuitive notion of the strength of information carried by an answer (such that, for example, ‘It’s borderline’ carries a stronger information than ‘It seems borderline’, even though they are logically independent), as nothing in our discussion will depend on the further fine details of the ordering. I will use ‘strong*’ and its like to denote it.

As I have already hinted at, it will be instructive to allow Vicky a wider range of answers than that determined by (E), while placing some additional constraints in order to capture what would suffice for the preservation of Vicky’s intellectual integrity in relation to the peculiarities of the setting of the forced march. I thus propose that we should require from forced-marched Vicky satisfaction of the following complex condition:

(HAPPY) Vicky gives a warranted answer for each question, and does so by:

(CLEAR) Having ‘Yes’ as the strongest* answer available to her for lectual integrity that we are not explicitly considering the possibility of answers expressing attitudes which differ in *degree* or in *kind* from flat-out judgement (such as, say, intermediate degrees of confidence and puzzlement respectively). As will be seen, all the decisive points in our dialectic can be more clumsily recast in such a way as to encompass this wider range of possible answers (one of whose limit cases is *deliberate silence*). Thanks to Crispin Wright for directing my attention to these possible alternative modes of answering.

patch #1, and ‘No’ as the strongest* answer available to her for patch #1,000;

(MAX) Giving always an answer which is maximally strong* among the answers warrantably available to her;

(KEEP) Giving the same answer as long as her state warrants such an answer and this is compatible with satisfaction of (MAX);

(COH) Not forming any unwarranted belief in the process as a consequence¹² of her answers (even if the belief were itself not expressed by any of her answers).

On the one hand, (CLEAR) is uncontroversial¹³ and (COH) is also hardly resistable: the integrity of an inquirer is arguably jeopardized as soon as she is forced to form an unwarranted belief in her attempt to give an answer, whether or not this belief is itself expressed by some of her answers ((COH) will play a crucial role in the dialectic of sections 6.4.3, 6.4.4, 6.5.1). On the other hand, (MAX) and (KEEP), despite their initial plausibility, may turn out to be too demanding—I will assume them for the time being, coming back in section 6.4.4 to the question of their correctness and of the consequences of their rejection. Also, clearly the spirit (if not the letter, depending on how the admissible/inadmissible distinction is finally spelled out) of (E) stands or falls with (HAPPY) (or possibly with one of its weaker versions without (MAX) or (KEEP) or either, depending again on how the admissible/inadmissible distinction is finally spelled out). We can thus hereafter focus on (HAPPY) and its weaker versions: in arguing that they do fall, we will have argued against (E) as well.

¹²In the *rational*, rather than *merely causal* sense of ‘consequence’. Thanks to Crispin Wright for pointing out the need for this qualification.

¹³Granted, it does ignore the possibility of Vicky’s having an answer e.g. for patch #1 which is logically stronger (and so stronger*) than ‘Yes’ (like ‘I know that patch #1 is red’), but this is just for ease of exposition—the arguments to follow can be more clumsily recast doing away with this simplification.

6.4.2 Sequentially Inconsistent Judgements

By (CLEAR) and (MAX), Vicky must then answer ‘Yes’ for patch #1. Presumably, she will return the same (we may assume equally warranted) verdict for patches #2, #3, #4 and quite a few others immediately following these, all at least very similar in colour to a clearly red patch. However, by (CLEAR), Vicky will not be warranted in returning the same verdict for patch #1,000. If she wants to achieve her aim, she must then return a different verdict at least for patch #1,000. *By classical logic*, there will be a *first* patch for which Vicky returns a verdict different from ‘Yes’. Indeed, since, by (CLEAR), she will not be warranted in returning for patch #1,000 any verdict which entails that patch #1,000 is red (as ‘I know that patch #1,000 is red’ would be), she must return at least for patch #1,000 a verdict which does not entail that the patch in question is red, and so, by classical logic, there will be a first patch for which Vicky returns such a verdict (hereafter, ‘different verdict’ and its like will usually be understood in this stronger sense).

Presumably, she will return the same (we may assume equally warranted) verdict she gave for patch #1,000 also for patches #999, #998, #997 and quite a few others immediately preceding these, all at least very similar in colour to a clearly orange patch. So let us suppose without loss of generality that the first patch for which Vicky returns a verdict different from ‘Yes’ is patch #401 (and let us suppose that what would intuitively be described as ‘the borderline cases of redness’ in effect start in the neighbourhood of patch #401, and end in the neighbourhood of patch #600). Thus, Vicky answered ‘Yes’ for patch #400 and *something different* for patch #401. The answer for patch #401 *need not* be ‘No’: Vicky might answer ‘There’s no fact of the matter’, or ‘I don’t know’, or ‘I don’t feel fully confident in saying ‘Yes’ ’ or in a myriad other ways.

For a start, let us however investigate what would happen if Vicky were to answer ‘No’ for patch #401. Then Vicky would in effect be judging patch

#400 to be red *and* be judging patch #401 not to be red (note the wide scope of ‘and’ here). But, for patches #400 and #401 as for any other two adjacent patches, Vicky cannot detect any difference in colour between them when considered in isolation from all other patches, and, more generally, cannot appeal (using perception or any other source of warrant) to any difference which she is warranted in taking to indicate a difference in redness.¹⁴ Hence, Vicky is certainly not warranted in judging that [patch #400 is red *and* patch #401 is not red] (note the narrow scope of ‘and’ here).¹⁵ Yet, Vicky could

¹⁴It is possible that there is a patch in the series that she could discriminate from patch #400 but not from patch #401, but why should she take this to indicate that patch #400 is red while patch #401 is not? (See Russell [1927], pp. 280–1; Russell [1950], pp. 104–5; Goodman [1951], pp. 196–200; Burns [1986]; Burns [1991], pp. 124–38 for relevant applications of this indirect-discrimination method and Dummett [1975a], pp. 321–3; Wright [1975], pp. 351–60; Williamson [1990], pp. 82–7 for criticisms thereof.)

¹⁵Sainsbury [1990], pp. 259–60; Sainsbury [1992], pp. 181–3 puts forth the example of an artist’s materials supplier who, having to arrange her red and yellow paints on just two shelves (marked ‘red’ and ‘yellow’ respectively), places as the least red paint on the ‘red’-marked shelf a paint *x* which is pairwise indiscriminable from the reddest paint *y* placed on the ‘yellow’-marked shelf (similar examples are given in Fara [2000], pp. 58–9). It is claimed that the supplier is warranted in doing so. I agree that, given her situation and purposes, the supplier is warranted in *placing* *x* on the ‘red’-marked shelf and *y* on the ‘yellow’-marked shelf, but reject as gratuitous the implication that she is also warranted in *judging* *x* to be red and *y* to be yellow. This is not to deny of course that ‘red’ and ‘yellow’ are in general context-dependent predicates or that, against their communal meaning, they might end up being used in the supplier’s shop in such a way that ‘*x* is red’ and ‘*y* is yellow’ are warranted for the supplier and her employees (just as an eccentric English mathematician might use in her own idiolect ‘prime number’ in such a way that ‘4 is a prime number’ is warranted for her). I thus regard the point correctly made by the example as rather irrelevant to Vicky’s situation.

Sainsbury [1992], pp. 181–3 puts also forth a more interesting example of a subject who is forced to judge a whole series of patches not sequentially, but in a random order, ending up with returning opposite verdicts for two adjacent pairwise indiscriminable patches. Unfortunately, an analogous example can be produced by using a precise predicate like ‘at least 20 ft tall’, where however the claim that the subject is warranted in returning opposite verdicts for, say, two adjacent pairwise indiscriminable trees is clearly preposterous. In view of this, I must confess that, without further supplementation, I don’t find any force in

still be warranted both in judging patch #400 to be red and in judging patch #401 not to be red if in this case warrant failed to collect over conjunction.

A suitably weak notion of warrant (as is most appropriate for our discussion) might very well entail failures of collection over conjunction, for, on a suitably weak construal, one is certainly warranted in believing, of each of the 1,000 possessors of a ticket of a fair lottery, that she won't win the lottery (assuming, of course, no collateral information), even though one is equally certainly not warranted in believing that none of them will win the lottery.¹⁶

the suggestion that, in Sainsbury's example, the subject is warranted in returning opposite verdicts for two adjacent pairwise indiscriminable patches. I'm indebted to Paula Milne and Mark Sainsbury for discussions of these and similar cases.

¹⁶Actually, assuming conjunction to be a fixed-arity 2ary operation, I don't accept the implicit reasoning here. For I take it to be very plausible that 'warranted' is vague in this context, in the sense that it is intuitively true that, if one is warranted in believing that $[[[\dots [x_0 \text{ won't win the lottery and } x_1 \text{ won't win the lottery}] \text{ and } x_2 \text{ won't win the lottery}] \text{ and } x_3 \text{ won't win the lottery}] \text{ and } x_4 \text{ won't win the lottery}] \dots \text{ and } x_i \text{ won't win the lottery}]$, one is also warranted in believing that $[[[[\dots [x_0 \text{ won't win the lottery and } x_1 \text{ won't win the lottery}] \text{ and } x_2 \text{ won't win the lottery}] \text{ and } x_3 \text{ won't win the lottery}] \text{ and } x_4 \text{ won't win the lottery}] \dots \text{ and } x_i \text{ won't win the lottery}] \text{ and } x_{i+1} \text{ won't win the lottery}]$ (with $x_0, x_1, x_2, \dots, x_{i+1}$ possessors of a ticket). This is just a tolerance principle as to of how many possessors one is warranted in believing that they won't win the lottery: one is certainly warranted in believing of one possessor that she won't win the lottery, one is certainly not warranted in believing of all the 1,000 possessors that they won't win the lottery, but of at most exactly how many possessors is one warranted in believing that they won't win the lottery?

Applying the naive theory of vagueness, the answer is that there is no such number, the tolerance principle being true. Since, in this case, a counterexample to collection will be something of the form 'One is warranted in believing that $[[[\dots [x_0 \text{ won't win the lottery and } x_1 \text{ won't win the lottery}] \text{ and } x_2 \text{ won't win the lottery}] \text{ and } x_3 \text{ won't win the lottery}] \text{ and } x_4 \text{ won't win the lottery}] \dots \text{ and } x_i \text{ won't win the lottery}]$, and one is warranted in believing that x_{i+1} won't win the lottery, but one is not warranted in believing that $[[[[\dots [x_0 \text{ won't win the lottery and } x_1 \text{ won't win the lottery}] \text{ and } x_2 \text{ won't win the lottery}] \text{ and } x_3 \text{ won't win the lottery}] \text{ and } x_4 \text{ won't win the lottery}] \dots \text{ and } x_i \text{ won't win the lottery}] \text{ and } x_{i+1} \text{ won't win the lottery}]$ ', the truth of any such counterexample would entail the falsity of the tolerance principle. Hence, the naive theory entails that, in this case at least, there is no such counterexample, and so that, in this case at least, collection is

But how could collection fail in this case? In the lottery case, what we may call the “*epistemic uncertainty*” of each conjunct (i.e. 1 minus its epistemic probability, see Edgington [1992], p. 193) is amplified in the conjunction, so as to reach at some point or other of the “conjoining” process a degree high enough as to undermine any warrant in the resulting conjunction; in the forced-march case, whatever epistemic uncertainty is present in each of the two discriminating judgements is amplified only once. Is Vicky’s warrant for each of the two discriminating judgements only ever so thin as to get lost at the very first stage of the conjoining process? This is implausible. For the series of patches might also change uniformly from big to small, and Vicky judge patch #401 not to be big: there is a strong intuition that, if she is warranted in judging patch #400 to be red and warranted in judging patch #401 not to be big, she is also warranted in judging that [patch #400 is red and patch #401 is not big].¹⁷ This shows that failure of collection, if such there is, cannot be attributed to the weakness of one’s warrant, in significant disanalogy to the lottery case.

It might be replied that the source of the failure of collection over conjunction is rather the fact that ‘Patch #401 is not red’ is *strongly* (negatively) probabilistically dependent on ‘Patch #400 is red’. Note that the appeal to strong probabilistic dependence at this stage is anyways mandated. For reflect that, in the lottery case, the epistemic uncertainty of each conjunction is

valid. The naive theory can agree though that the case does present a counterexample to collection once conjunction is interpreted as a variable-arity operation (in particular, the counterexample would target collection over the 1,000ary conjunction saying that [x_0 won’t win the lottery and x_1 won’t win the lottery and x_2 won’t win the lottery and x_3 won’t win the lottery and x_4 won’t win the lottery... and $x_{1,000}$ won’t win the lottery]).

¹⁷This much is of course perfectly compatible with the relevant epistemic probabilities obeying the classical laws of probability, and so with Vicky’s epistemic probability for ‘Patch #400 is red and patch #401 is not big’ being the product of her epistemic probabilities for ‘Patch #400 is red’ and ‘Patch #401 is not big’ (Schiffer [1998]; Schiffer [2000]; Schiffer [2003], pp. 178–237 uses related but more problematic considerations to argue for the revision of the classical laws of probability in the presence of vagueness-related partial beliefs).

only ever so slightly higher than the maximum of the epistemic uncertainties of each conjuncts, while, in the forced-march case, as has just been argued, the epistemic uncertainty of ‘Patch #400 is red and patch #401 is not red’ is very close to 1 (assuming e.g. that there are 200 equally reasonable candidates for being the cut-off point for redness in the series, and assuming a plausible principle of indifference, it will be .995), and so considerably higher than the maximum of the epistemic uncertainties of ‘Patch #400 is red’ and ‘Patch #401 is not red’. In order for failure of collection to be sustained, this would require, in significant disanalogy to the lottery case (where only a very weak negative probabilistic dependence obtains), that ‘Patch #400 is red’ and ‘Patch #401 is not red’ be strongly negatively probabilistically dependent.

I am willing to concede here that it is plausible to set the conditional epistemic probability of ‘Patch #401 is not red’ on ‘Patch #400 is red’ significantly lower than the epistemic probability of ‘Patch #401 is not red’. A simple model of this situation can be given by letting there to be 200 equally reasonable candidates for being the cut-off point for redness in the series (say, from patch #400 to patch #599), and setting to .5 the epistemic probability for each such patch to be red. Then, for example, the epistemic probability of ‘Patch #401 is not red’ is .5, while, assuming the same principle of indifference as in the last paragraph, the conditional epistemic probability of ‘Patch #401 is not red’ on ‘Patch #400 is red’ can be as low as .01 (thus showing the strong negative probabilistic dependence of the former on the latter), and so the epistemic probability of ‘Patch #400 is red and patch #401 is not red’ can fall well below .25 (namely, as low as .005).

To be sure, one could contest the low assignment of conditional epistemic probability; one could also contest both the model’s faithfulness to Vicky’s situation and the claim that what is modelled is a failure of warrant to collect over conjunction. While I think that these are legitimate worries and need be addressed, what I would like to question here is rather whether there are cases where φ and ψ are such that the conditional probability of the negation

of the latter on the former is very high, and yet one is both warranted for φ and warranted for ψ without being warranted for ' φ and ψ '. Consider the following case:

JOHANN AND JOSEPH Johann is an Italian citizen grown up in South Tyrol, always speaking German and never learning Italian. Meeting Johann at the airport, Joseph can gather some evidence that Johann is Italian by noticing that he's flying back to Venice after what have apparently been his holidays, and also gather some evidence that he doesn't speak Italian by seeing him reading a German newspaper. Pointing at Johann, Joseph's conditional epistemic probability for 'That man speaks Italian' on 'That man is Italian' is very high, even though Joseph's epistemic probabilities for both 'That man is Italian' and 'That man doesn't speak Italian' are not too low. Presumably, the flimsy evidence available to Joseph is not sufficient to warrant him in believing to be in the presence of the rather bizarre case of an Italian who doesn't speak Italian. But is Joseph nevertheless warranted both in believing that Johann is Italian and in believing that he doesn't speak Italian? It seems clear that he isn't, even though, given Joseph's state of information, the two new pieces of evidence may well defeat both his antecedent warrant for believing Johann not to be Italian and his antecedent warrant for believing Johann to speak Italian.

Consideration of this and similar cases invites some inductive scepticism with regard to the hypothesized configuration of epistemic probabilities and warrant. Note also that, even if the hypothesis should prove to be correct for some other (presumably, rather exotic) cases, it would still have to be argued in addition that the forced-march case can be assimilated to them in the relevant respects.

These are however relatively minor niceties. For the simple and yet damning point is that Vicky's *pair of attitudes*—belief that patch #400 is red, belief that patch #401 is not red—draws a sharp distinction between patch #400

and patch #401 for which Vicky has no warrant: for all Vicky can tell, an analogous pattern of attitudes placing the sharp boundary in question rather between, say, patch #401 and patch #402 would have been just as good (or just as bad). It is important to note that the present point requires the legitimacy of attaching the epistemic property of warrant (or lack thereof) to a *pattern* of attitudes rather than to a *single* attitude: I believe that we have a clear enough intuitive grasp of this wider range of attributions and, in the following, will continue to work under this plausible assumption and a broader understanding of (HAPPY) which makes its satisfaction sensitive also to these issues.

Of course, there may be cases where arguably a pattern of attitudes is not only such that each of the attitudes is warranted but is also itself warranted, even though the single attitude towards the conjunctive content generated by conjoining the contents of all the attitudes (see fn 16) is still not warranted—lottery-paradox cases are arguably such. In this respect, our case is strikingly different, as the lack of warrant is already present with the pattern of differential attitudes towards adjacent patches, even without considering the further possible single attitude arising from the conjunction of the contents of these two attitudes. Moreover, note that the lack of warrant only immediately attaches to the pattern of differential attitudes towards adjacent patches rather than to either attitude considered singularly. For, at this point of the dialectic, we can grant not only that Vicky both may well be warranted in judging patch #400 to be red and may well be warranted in judging patch #401 not to be red (note here the wide scope of ‘and’ with respect to ‘may’), but also that she may well be both warranted in judging patch #400 to be red and warranted in judging patch #401 not to be red (note here the narrow scope of ‘and’ with respect to ‘may’ and its wide scope with respect to ‘warranted’)—we only have to insist that she is not warranted in both judging patch #400 to be red and judging patch #401 not to be red (note here the narrow scope of ‘and’ with respect to ‘warranted’).

Before bringing this particular discussion to an end, I want to mention

briefly an alternative line of attack to the use of collection of warrant in this dialectic. The attack takes inspiration from the contextualist idea that, when Vicky returns a different verdict for patch #401, the extension of ‘red’ as used by her changes in such a way that neither patch #401 *nor* patch #400 belong to it, while, when Vicky returned a verdict for patch #400, both patch #400 *and* patch #401 belonged to it (see Raffman [1994]; Raffman [1996]; Soames [1999], pp. 203–27; Fara [2000]; Shapiro [2003]; Shapiro [2006], pp. 1–44 for specific implementations of this idea and Robertson [2000]; Keefe [2003]; Keefe [2007]; Heck [2003], pp. 118–20; Stanley [2003] for some influential criticisms). If this idea is correct, then Vicky might well be warranted in both her judgements without being warranted in identifying any relevant sharp boundary. Generally speaking, this attack inherits many of the flaws of contextualist approaches to vagueness, whose discussion lies outside the scope of this essay. Here, I will rest content with two critical observations specifically pertaining to our problem. Firstly, satisfaction of (HAPPY) achieved because of the psychological or pragmatic or semantic impossibility of keeping fixed one’s words’ extensions is just another (not so subtle) version of the strategy of changing the subject (already encountered in other versions in sections 6.3.2, 6.4.1) and as such rather irrelevant to our guiding question concerning the preservation of Vicky’s intellectual integrity. Secondly, no matter what its (dubious) merits are for forced marches using vague predicates, the strategy is a non-starter for forced marches using precise predicates (as those discussed in section 6.4.5) and so is at best radically incomplete.¹⁸

The conclusion is then that, if collection of warrant holds (as seems plausible in this case), since Vicky is not warranted in judging that [patch #400 is red and patch #401 is not red], it is not the case that Vicky is both warranted in judging patch #400 to be red and warranted in judging patch #401 not to be red (even though, for all that has been said, she might well be warranted in *exactly one* of these judgements, and be warranted in the

¹⁸Thanks to Mark Sainsbury and Crispin Wright for pointing out to me the need to mention this contextualist move.

general judgement that there is a cut-off point for redness in the series)—if collection of warrant does not hold, we are still stuck with the fact that it is not the case that Vicky is warranted both in judging patch #400 to be red and in judging patch #401 not to be red. Hence, either way, Vicky's answer for patch #401 must be different from 'No'.

Let us now first consider the case where her answer is still inconsistent with patch #401's being red (such would plausibly be if it were e.g. 'There's no fact of the matter', 'That's not determinate', 'That's not true' etc.). In that case, the previous consideration concerning Vicky's inability to detect any difference in colour between patches #400 and #401 when considered in isolation from all other patches, and, more generally, Vicky's inability to appeal to any difference which she is warranted in taking to indicate a difference in redness would still apply to a judgement that [patch #400 is red and patch #401 is not determinately red] (using 'not determinately red' as a catch-all phrase for any of the specific answers under consideration). For, since being red is inconsistent with not being determinately red, they mark among things a genuine (albeit more exotic) *difference* in redness just as being red and not being red do: that is, just as, in matters of redness, being not red *rules out* being red, so, in matters of redness, not being determinately red also *rules out* being red (even though it need not imply not being red). And it seems that Vicky is no better off in detecting a difference of *this* kind between patch #400 and patch #401 than she was in detecting a red/non-red difference between them.

6.4.3 Sequentially Consistent Judgements

We now turn to consider the case where Vicky's answer is consistent with patch #401's being red (such would plausibly be if it were e.g. 'I don't know', or 'I don't feel fully confident in saying 'Yes'', or 'I don't know that I know' etc.). Notice first that the specific problem which used to arise for the previous kind of answer does not arise for this new kind of answer. For

the problem used to arise because, in giving inconsistent answers for patch #400 and patch #401, Vicky was committed to being warranted in drawing a difference in redness for which she seemed to have no warrant. But, clearly, no difference in redness between the two patches need be drawn in judging that one is red and that the other is not known to be red: the judgement is still compatible with both patches being just plain red.

Nevertheless, suppose that Vicky would have been warranted in judging patch #401 to be red: then, in not answering ‘Yes’ to the question as to whether patch #401 is red, she would be violating (KEEP) (since the judgement in question is a judgement to the effect that a certain patch is red, there is no question of (MAX) overriding (KEEP) in this case). Hence, if Vicky satisfies (HAPPY), she is not warranted in judging patch #401 to be red. This simple piece of reasoning is however available to Vicky: she is thus warranted in believing that, if she does not answer ‘Yes’ to the question as to whether patch #401 is red, if she satisfies (HAPPY), she is not warranted in judging patch #401 to be red. Moreover, presumably, there should be no bar to her being warranted in believing that she has not answered ‘Yes’ to the question as to whether patch #401 is red.¹⁹ If we make the additional assumption that Vicky must be warranted in believing that she satisfies (HAPPY), it follows, by what seems to be two ungainsayable applications of closure of warrant under logical consequence, that she is warranted in believing that she is not warranted in judging patch #401 to be red.

Unfortunately, Vicky answers ‘Yes’ to the question as to whether patch #400 is red. We can thus use an analogous argument to show that she is warranted in believing that she is warranted in judging patch #400 to be red. Since Vicky answers ‘Yes’ to the question as to whether patch #400 is red, if she satisfies (HAPPY), she is warranted in judging patch #400 to be red. This simple piece of reasoning is available to Vicky: she is thus warranted in believing that, if she answers ‘Yes’ to the question as to whether

¹⁹Of course, even if the *content* of her answer is consistent with the content of the answer ‘Yes’, her *action* is inconsistent with the action of answering ‘Yes’.

patch #400 is red, if she satisfies (HAPPY), she is warranted in judging patch #400 to be red. Moreover, presumably, there should be no bar to her being warranted in believing that she has answered ‘Yes’ to the question as to whether patch #400 is red. Given the same additional assumption that Vicky must be warranted in believing that she satisfies (HAPPY), it follows, by what seems to be two ungainsayable applications of closure of warrant under logical consequence, that she is warranted in believing that she is warranted in judging patch #400 to be red.

Therefore, even if Vicky’s answer is consistent with patch #401’s being red, her not answering ‘Yes’ to the question as to whether patch #401 is red still entails, together with the other assumptions made explicit in the last two paragraphs, that [[Vicky is warranted in believing that she is warranted in judging patch #400 to be red] and [Vicky is warranted in believing that she is not warranted in judging patch #401 to be red]]. By collection of warrant over conjunction, Vicky would be warranted in believing that [she is warranted in judging patch #400 to be red and not warranted in judging patch #401 to be red]. However, ‘Vicky is warranted in judging ξ to be red’ seems to be no less vague than ‘ ξ is red’: which is the last patch that Vicky is warranted in judging to be red? This vagueness reflects itself in the fact that, for patches #400 and #401 as for any other two adjacent patches, Vicky cannot detect any difference in warranting a ‘Yes’-judgement between them when considered in isolation from all other patches, and, more generally, cannot appeal (using perception or any other source of warrant) to any difference which she is warranted in taking to indicate a difference in warranting a ‘Yes’-judgement (see fn 14). Hence, Vicky is certainly not warranted in judging that [she is warranted in judging patch #400 to be red and not warranted in judging patch #401 to be red].

The foregoing argument makes use of the crucial additional assumption that Vicky must be warranted in believing that she satisfies (HAPPY). This assumption is not at all uncontroversial, as it requires a mild iteration principle for warrant: Vicky cannot just be warranted in her first-order judgements

about the patches, but must also be warranted in her second-order beliefs that she is so (first-order) warranted. Indeed, whatever norm N we should like to impose on an inquiry in order to preserve an inquirer's intellectual integrity, N will have to entail that the inquirer gives warranted answers to the innocent questions posed to her: an analogous assumption to the effect that the inquirer must be warranted in believing that she satisfies N would result in the fact that the inquirer cannot just be warranted in her first-order judgements about whatever her inquiry is about, but must also be warranted in her second-order beliefs that she is so (first-order) warranted. One who is doubtful about the general truth of the characteristic **S4** axiom for warrant (if one is warranted in believing that P , then one is warranted in believing that one is warranted in believing that P) may well question at this point the legitimacy of this kind of assumption.

In addressing this worry, I would like to point out first that I actually regard the present dialectic as an exemplification of the problematicity of attaching significant normative force to a certain epistemic property of being E while denying the general truth of the characteristic **S4** axiom for that property. For such a combination has to envisage situations in which an agent meets a norm for $\mathcal{A}\text{ing}$ ²⁰ that P in virtue of her being E with respect to the proposition that P , but fails to be E with respect to the proposition that [she is E with respect to the proposition that P]. Such an agent would thus fail to meet a norm for $\mathcal{A}\text{ing}$ that [she is E with respect to the proposition that P], and, since being E is at least a necessary condition for correct $\mathcal{A}\text{ing}$, by closure of being E under logical consequence she would presumably fail to meet a norm for $\mathcal{A}\text{ing}$ that she correctly $\mathcal{A}s$ that P .

For example, knowledge might be proposed as a norm for asserting and the general truth of the characteristic **S4** axiom for knowledge be denied (as in Williamson [2000b]). Such a combination has to envisage situations in which an agent meets the knowledge norm for asserting that it's cold outside

²⁰Roughly, for every attitude of $\mathcal{A}\text{ing}$, a *norm* for $\mathcal{A}\text{ing}$ is a necessary condition for an event of $\mathcal{A}\text{ing}$ to be correct.

in virtue of her knowing that it's cold outside, but fails to know that she knows that it's cold outside. Such an agent would thus fail to meet a norm for asserting that she knows that it's cold outside, and, since knowledge is at least a necessary condition for correct asserting, by closure of knowledge under logical consequence she would presumably fail to meet the knowledge norm for asserting that she is correctly asserting that it's cold outside. Assuming further that the agent meets whatever other necessary condition is in place for correctly asserting that it's cold outside, we would end up in a situation where the norms of assertion allow an agent to assert that it's cold outside, but forbid her to defend herself in the face of challenges to the correctness of her assertion. This, I submit, is not assertion as we know and love it: either one keeps silent in the first place, or one should be allowed to go on and defend oneself from charges of incorrectness—to have to stop midway would not just be extremely bizarre, but would also be hardly compatible with the preservation of one's intellectual integrity, running against any minimal ideals of self-awareness and coherence in inquiry.

Worries from the alleged problematicity of the characteristic **S4** axiom for warrant may also be allayed by remarking that what the additional assumption in question requires is only that Vicky be warranted in believing that she satisfies (HAPPY) as far as her *first-order* judgements go: it does not launch Vicky in a dangerous climb up the hierarchy of warrant by requiring that she be warranted in believing that she satisfies (HAPPY) as far as *any* of her first-order or higher-order judgements go. This mild iteration of warrant is perfectly compatible with the general falsity of the characteristic **S4** axiom for warrant (and compatible with the results of the most challenging arguments against the axiom, like that advanced in Williamson [1992]; Williamson [1996]; Williamson [2000b], pp. 93–134). Indeed, by (COH), if Vicky cannot warrantably believe that she satisfies (HAPPY), she is not even allowed to believe that she satisfies (HAPPY). But how could this belief be dispensed with? How could she try to conform to (HAPPY) without even believing that she is so conforming?

To clinch matters, note that there should be no objection to Vicky's being herself a believer in the failure of iteration, if that's the truth on these matters. After all, failure of iteration is now supposed to be the reason why Vicky should not believe that she is conforming to (HAPPY). Vicky would then have to believe that, for some of her first-order judgements to the effect that P , she is warranted in judging that P and not warranted in believing that she is warranted in judging that P . Note that this by itself is not at all (Moore) paradoxical: it is just an instance of the perfectly appropriate (for us) quantified belief that, for some P , P and one is not warranted in believing that P (note the wide scope of 'belief' here).

The problem arises rather by considering that it would seem that Vicky can also be warranted in believing what some of the witnesses of the quantified belief are: she is warranted in believing that she switches after patch #400, and it seems that she can also be warranted in believing that, if iteration fails at all, it will fail for the patch after which she switches (after all, failure of iteration at patch #400 is now supposed to be the reason why Vicky should not believe that she is conforming to (HAPPY) for patch #400). Unfortunately, assuming mild closure-under-logical-consequence principles for warrant, warrant in believing in a particular failure of iteration of warrant is not only Moore paradoxical (having the general form that a conjunction saying that [P and the proposition that P lacks a certain doxastic or epistemic property of being E] is itself E), but it also entails the non-failure of the iteration and, assuming (HW), is indeed impossible. For a subject s 's warrant in believing that [s is warranted in believing that P and s is not warranted in believing that s is warranted in believing that P] entails, by a mild principle of closure of warrant under logical consequence, both that [s is warranted in believing that s is warranted in believing that P] and that [s is warranted in believing that s is not warranted in believing that s is warranted in believing that P].

6.4.4 Unwise Chrysippus

It might be proposed at this point that the trouble with the forced march resides in (MAX)—which, despite its initial plausibility, should therefore be rejected as too demanding. Rejecting (MAX) (but keeping (KEEP)), Vicky could start by answering e.g. ‘It’s not definitely orange’ for patch #1 and then switch to ‘It’s not definitely red’ at patch #401.²¹ Importantly, it may seem that a rejection of (MAX) could be adopted in a principled way by noticing that, under some natural assumptions, what (MAX) really amounts to is the *conformity* condition:

(CONF) Vicky must always conform her judgements to whatever positive warrant is available to her.

Together with the following principle linking *conforming* and *warrant*:

(CW) Conforming to its being the case that P (or, at least, conforming to its being the case that there is warrant for one of Vicky’s judgements) requires warrant in judging that P (warrant in judging that there is warrant for one of Vicky’s judgements),

²¹Note in passing that the converse strategy of keeping (MAX) and rejecting (KEEP) does not seem even to get off the ground. For, by (CLEAR) and (MAX), Vicky must answer ‘Yes’ for patch #1 and ‘No’ for patch #1,000. By classical logic, there will be a first patch # i for which Vicky returns a verdict different from ‘Yes’. However, given the ordering of informational strength, when she stops answering ‘Yes’ for patch # i , she can only do so by replying ‘No’. For suppose that she replies anything weaker* than ‘No’. Then, this will also be weaker* than ‘Yes’. If she satisfies (HAPPY) (in particular, (MAX)), this can only be so if she is not warranted in judging patch # i to be red. The main line of argument of section 6.4.3 could then be applied to obtain the conclusion that she unwarrantedly believes that she is warranted in judging patch # $i - 1$ to be red and not warranted in judging patch # i to be red. If however Vicky replies ‘No’, then the main line of argument of section 6.4.2 could be applied to obtain the conclusion that she unwarrantedly believes patch # $i - 1$ to be red and patch # i not to be red.

(CONF) would then amount to a local imposition of the characteristic **S4** axiom for warrant, which may cause the usual worries.

I hope to have made a case in section 6.4.3 that, even if this result should obtain, such worries are likely to be misplaced. What I want to point out here is that (CW) seems to presuppose an excessively intellectualist interpretation of what conforming to warrant must amount to. For presumably warrant for judging that P (where the proposition that P is an object of one of Vicky's judgements) *requires* a reliable disposition to judge that P , at least under appropriate circumstances (at the very least, it should certainly be *compatible* with such a disposition, and we can imagine Vicky to be a being where warrant and disposition do co-exist), and the existence of such a disposition would seem sufficient for one to conform to one's warrant. One only needs to achieve an adequate matching between one's warrant in judging that P (a warrant for i th-order judgements) and one's judgements that P (i th-order judgements)—one need not achieve an adequate matching between one's warrant in judging that P (a warrant for i th-order judgements) and one's judgements that one is warranted in judging that P ($i + 1$ th-order judgements). However, I will leave open whether there exists a principled route to rejection of (MAX), and pursue rather the question of what its effects would be.

Indeed, even if no flaw should lie in the motivation for rejecting (MAX), the main trouble would still reside rather in its little effects in saving Vicky from her predicament. We must first note that matters cannot be left at a blanket rejection of (MAX), for then (HAPPY) could be satisfied by Vicky by just answering 'It is not the case that it is both red and not red' throughout. Accepting only trivialities in the face of overwhelming positive or negative warrant is a parody of inquiry—such a satisfaction of (HAPPY) achieved by changing the subject would amount to a Pyrrhic victory. Clearly, some *norm* enjoining straightforwardly positive or negative judgements for at least some cases must remain in place even after (MAX) has been discarded. The norm should probably enjoin a straightforwardly positive judgement for patch #1,

which implies rejection of (KEEP) along with (MAX) (otherwise the main line of argument of section 6.4.3 could once again be applied). Again, I will leave open whether there exists a principled route to rejection of (KEEP) (this would probably involve a dialectic analogous to the one we have just broached for rejection of (MAX)), and pursue rather the question of what its effects would be.

Within the framework of a joint rejection of (MAX) and (KEEP), the following more satisfactory strategy would then be permissible. Vicky starts by answering ‘Yes’, and does so for the first few patches, until she reaches patches which come both decently after patch #1 and comfortably before the last patch for which ‘Yes’ would be a warranted answer. Let us suppose without loss of generality that the first patch for which Vicky returns a verdict different from ‘Yes’ is patch #351. Vicky must of course provide an alternative answer. Something along the lines of ‘It’s either red or borderline red’ would do: for Vicky could then switch (in a second violation of (MAX) and (KEEP)) to ‘It’s either orange or borderline red’ at, say, patch #501 (and then finally switch to ‘No’ at, say, patch #651). Something along these lines seems indeed to have been the strategy advocated by the Stoics (see Barnes [1982]; Burnyeat [1982]; Mignucci [1993] for the sources and Williamson [1994], pp. 8–27 for a more theoretical and highly sympathetic exposition).

But what must Vicky believe when she switches at patch #351? As anticipated, she must be complying with a norm N which, in addition to forbidding assent for patches too close to the last patch for which assent is warranted, also enjoins assent for patches very close to patch #1. Vicky must then believe that her switch complies with N . It is actually very natural to interpret N as enjoining to assent *as long as possible* (that is, just before one’s switch would be too close to the last patch for which assent is warranted): epistemic norms do seem to be *maximizing* norms which enjoin belief in all propositions of a certain kind. Such an interpretation of N would however clearly make impracticable the strategy under consideration, and so, without

committing ourselves to the ultimate admissibility of a weaker reading of N , we will hereafter assume that N only:

- Enjoins assent for patches very far from the last patch for which assent is warranted;
- Allows for assent and for something different than assent for patches not very far from but also not too close to the last patch for which assent is warranted;
- Forbids assent for patches too close to the last patch for which assent is warranted.

Of course, Vicky should not believe that her switch occurs at the *first* or *last* patch for which a switch would comply with N , for ‘ ξ is such that a switch at it would comply with N ’ (henceforth, ‘ ξ is switchable’) is just as vague as ‘ ξ is red’ is, so that all the problems we have been reviewing would re-occur for such a switch.

A similar problem would occur if Vicky were to believe that patch #351 is the *best* patch to switch at, for ‘ ξ is the best patch to switch at’ is similarly vague (at least in the sense that we and Vicky are now not able to provide a warranted identification of its cut-off point, which is the relevant sense in our epistemological setting). It is actually not easy to make sense of how Vicky can switch at patch #351 without believing it to be the best patch to switch at: if, say, patch #350 is believed by her to be just as good a patch to switch at as patch #351, why does she switch at patch #351 rather than at patch #350?

It is at least unclear that the lack of any epistemic (or practical) reason for such a “choice” could be counterbalanced by the kind of practical considerations usually appealed to e.g. in the case of Buridan’s ass. What needs to be rationalized is not any old action, but a change in *judgement*. Vicky does have an epistemic (and practical) reason for bringing it about

that, for some switchable patch x , she switches at x , but for no switchable patch x does she have an epistemic (or practical) reason to bring it about that she switches at x . It would seem however that, in such a case where judgements are concerned, nothing short of an epistemic (or at least practical) reason for switching at x would warrant Vicky in switching at x , and so that for no switchable patch x is Vicky warranted in switching at x . Of course, elaborating on a theme of section 6.4.2, what would be unwarranted are neither Vicky's answer for patch #350 nor her answer for patch #351, but her switching from one to the other (the pattern of her answers), just as, when presented with two trees of exactly the same height, one's pattern of answers 'That is taller than 20 ft' (pointing at one tree) and 'That is between 30 and 10 ft tall' (pointing at the other tree) would be unwarranted, even though each single answer may well be warranted.

Be that as it may, the most resilient form of the problem would still remain. Patches #1 and #1,000 are not only not switchable, but also such that Vicky is not warranted in believing them to be switchable, and so, by classical logic, there will be a first patch # i and a last patch # j [$1 < i \leq 351$, $351 \leq j < 1,000$] warrantably believable by Vicky to be switchable.²² Moreover, the strategy we are considering requires Vicky to believe, at least of the patch (#351) she switches at, that that patch is switchable. For one has to believe that a patch is switchable in order for one rationally to decide to switch at it, just as one has to believe that a chair is movable in order for one rationally to decide to move it. Vicky can thus only start to believe that a patch # k is switchable if $k \geq i$, and only stop to believe that a patch # l is switchable if $l \leq j$. Given (COH), all these beliefs about switchability must also be warranted. Were Vicky to try to comply with appropriate versions

²²Hereafter, for simplicity's sake, we will assume classical logic also for vague predicates like 'ξ is warrantably believable by Vicky to be switchable'. Such an assumption plays no essential role in the argument and could be replaced by appeal to vague sets or the like. For example, instead of talking about patch # i and patch # j , we could have appealed to the (vague) set of patches warrantably believable by Vicky to be switchable, set to which neither patch #1 nor patch #1,000 belongs.

of (MAX) and (KEEP) concerning switchability, she would of course incur problems analogous to those we have been reviewing in sections 6.4.2, 6.4.3.

A strategy analogous to the one just exposed must then be deployed, allowing Vicky to switch to ‘switchable’ (“*positive*” switch) comfortably after patch $\#i$ (let us call the ordinal number of the first such patch ‘ $\text{mar}^{pos}(\text{switch})$ ’) and switch from ‘switchable’ (“*negative*” switch) comfortably before patch $\#j$ (let us call the ordinal number of the first such patch ‘ $\text{mar}^{neg}(\text{switch})$ ’).

Definition 6.4.1. A predicate Φ is *convex* iff, for some m_0, m_1 [$0 < m_0 < m_1 < 1,000$], the extension of Φ is $\{\text{patch } \#n : m_0 \leq n \leq m_1\}$.

Definition 6.4.2. For Φ convex:

- $\text{mar}^{pos}(\Phi)$:= the first m such that patch $\#m$ comes comfortably after the first patch warrantably believable by Vicky to be in the extension of Φ ;
- $\text{mar}^{neg}(\Phi)$:= the last m such that patch $\#m$ comes comfortably before the last patch warrantably believable by Vicky to be in the extension of Φ .

Hierarchy starts looming here, and we will call ‘switchable²’ all and only those patches that lie between patch $\#\text{mar}^{pos}(\text{switch})$ and patch $\#\text{mar}^{neg}(\text{switch})$. For every i , ‘switchable ^{i} ’ (and its like, e.g. ‘switches ^{i} ’), as well as positive and negative versions thereof, can be analogously defined by recursion:

Definition 6.4.3.

- patch $\#n$ is *switchable*¹ := patch $\#n$ is switchable;
- patch $\#n$ is *switchable* ^{$m+1$} := $\text{mar}^{pos}(\text{switch}^m) \leq n \leq \text{mar}^{neg}(\text{switch}^m)$.

Definition 6.4.4.

- patch $\#n$ is *positively switchable*¹ := patch $\#n$ is patch $\#1$;
- patch $\#n$ is *positively switchable* ^{$m+1$} := $\text{mar}^{\text{pos}}(\text{'switch'}^m) \leq n$.

Definition 6.4.5.

- patch $\#n$ is *negatively switchable*¹ := patch $\#n$ is not too close to the last patch for which assent to ‘Is this patch red?’ is warranted;
- patch $\#n$ is *negatively switchable* ^{$m+1$} := $n \leq \text{mar}^{\text{neg}}(\text{'switch'}^m)$.

Theorem 6.4.1. *For every m [$m : m > 1$], n , patch $\#n$ is switchable ^{m} iff patch $\#n$ is both positively and negatively switchable ^{m} .*

Proof. Immediate from definitions 6.4.3, 6.4.4, 6.4.5.

□

Theorem 6.4.2. *For every m , $\{x : x \text{ is switchable}^{m+1}\} \subseteq \{x : x \text{ is switchable}^m\}$, and, if $\{x : x \text{ is switchable}^m\} \neq \emptyset$, $\{x : x \text{ is switchable}^{m+1}\} \subset \{x : x \text{ is switchable}^m\}$.*

Proof. Immediate from definition 6.4.3.

□

Theorem 6.4.3. *For some m , $\{x : x \text{ is switchable}^m\} = \emptyset$.*

Proof. Immediate from theorem 6.4.2 and the finitude of $\{x : x \text{ is switchable}\}$.

□

Theorem 6.4.4. *For every m , patches $\#1$, $\#1,000$ are not switchable ^{m} .*

Proof. By induction.

- By construction, patches #1, #1,000 are not switchable¹;
- By theorem 6.4.2, for every m , $\{x : x \text{ is switchable}^{m+1}\} \subseteq \{x : x \text{ is switchable}^m\}$. Since, by the induction hypothesis, patches #1, #1,000 are not switchable ^{m} , they are not switchable ^{$m+1$} either.

□

Now, reflect that, generalizing the previous point concerning the requirement of Vicky to believe, at least of the patch (#351) she switches at, that that patch is switchable, we have the *belief* requirement:

(BEL) It is required of Vicky that:

- If she switches at a certain patch, she believe that patch to be switchable;
- For every m [$m : m > 1$], if she positively (negatively) switches ^{m} at a certain patch, she believe that patch to be positively (negatively) switchable ^{m} .

Adding the *monotonicity* principle:

(MON) For every m [$m : m > 1$], n_0 , if Vicky believes that patch # n_0 is positively (negatively) switchable ^{m} , she believes, for every n_1 [$n_1 : n_1 \geq n_0$] (for every n_1 [$n_1 : n_1 \leq n_0$]), that patch # n_1 is positively (negatively) switchable ^{m} ,

we obtain:

Theorem 6.4.5. *For every m , n , if Vicky believes patch # n to be switchable ^{m} , she believes it to be switchable ^{$m+1$} .*

Proof. Suppose that Vicky believes patch $\#n$ to be switchable m . Then, by theorem 6.4.4, for some n_0, n_1 [$n_0 \leq n \leq n_1$], she must positively switch $^{m+1}$ at patch $\#n_0$ and negatively switch $^{m+1}$ at patch $\#n_1$. By (BEL), Vicky believes patch n_0 to be positively switchable $^{m+1}$ and patch n_1 to be negatively switchable $^{m+1}$. By (MON), she believes patch $\#n$ to be both positively switchable $^{m+1}$ and negatively switchable $^{m+1}$, and so, by her knowledge of theorem 6.4.1, she believes it to be switchable $^{m+1}$.

□

Theorem 6.4.6. *For every m , Vicky believes patch $\#351$ to be switchable m .*

Proof. By induction.

- By assumption, Vicky believes patch $\#351$ to be switchable 1 .
- By the induction hypothesis, Vicky believes patch $\#351$ to be switchable m , and so, by theorem 6.4.5, she believes it to be switchable $^{m+1}$.

□

It then follows in particular from theorems 6.4.3, 6.4.6 that, for some m , Vicky believes patch $\#351$ to be switchable m without patch $\#351$'s being switchable $^{m+1}$, contrary to what the strategy under consideration requires. Whatever its other merits may be, we can thus conclude that it cannot be implemented in the general way required to save Vicky from her predicament.

It is important to stress that the crucial use made by the above argument of (BEL) (with the background assumption of (COH)) by no means amounts to some transparency requirement which would appear problematic from the perspective of a rejection of the characteristic **S4** axiom for warrant (setting aside now the question of how much mileage there is in such a rejection for the issues at hand).

Firstly, inspection of the definition of ‘ ξ is switchable^{*i*}’ reveals that the notion need not involve any epistemic feature problematic for transparency: it is of course *stronger* than the notion expressed by ‘It can be warrantably believed that ξ is switchable^{*i-1*}’, but it also need *not* be *as weak as* any of the notions expressed by ‘It can be warrantably believed that it can be warrantably believed that it can be warrantably believed...that it can be warrantably believed that ξ is switchable^{*i-1*}’, for any iteration of the ‘It can be warrantably believed that’-operator. Hence, (BEL) (with the background assumption of (COH)) need not require any form of transparency of warrant. Moreover, the definition of ‘ ξ is switchable^{*i*}’ only requires a patch to be within a certain range (to be specified), and the specification of such a range need only appeal to numerical values which are known to be *sufficient* for creating the necessary margins of safety—crucially, they need not be *necessary* for doing so. Some such values can certainly be specified, and once ‘ ξ is switchable^{*i*}’ is made precise in this way, (BEL) (with the background assumption of (COH)) should be wholly unproblematic.

Secondly, (BEL) (with the background assumption of (COH)) does not require in any case that, in general, if a patch is switchable^{*i*}, Vicky is warranted in believing that it is so: while it does require Vicky to believe that a patch is positively (negatively) switchable^{*i*} if she decides positively (negatively) to switch^{*i*} at that patch, it does not require her to track the full range of switchability^{*i*}.

Thirdly, even if fault should be found with (BEL) on general grounds, it is hard to see how it could be dispensed with by the strategy under consideration: for the strategy must certainly allow (indeed—it would seem—require) that Vicky at least believes to be carrying out the strategy, and (BEL) amounts to no more than this condition. Saying that, in order to preserve her intellectual integrity, Vicky should carry out the strategy but not allowing her at least to believe to be carrying it out would once again seem to run against any minimal ideals of self-awareness and coherence in inquiry.

Fourthly, a weaker version of (BEL) could be formulated as follows:

(BEL⁻) It is required of Vicky that:

- If she switches at a certain patch, she believe *that some* patch is switchable;
- For every m [$m : m > 1$], if she positively (negatively) switches ^{m} at a certain patch, she believe *that some* patch is positively (negatively) switchable ^{m} ,

requiring only *general de dicto* beliefs about patches rather than *de re* beliefs about *specific* patches (a version intermediate in strength between (BEL) and (BEL⁻) would require *de re* beliefs about *some* patches or others). (BEL⁻) would seem unnegotiable, and yet a structurally similar argument would still go through (I leave the details to the reader).

6.4.5 The Enforcement of Classical Logic: Transparency and Inexact Knowledge

It may at this point be objected that all these ruminations crucially rely on the assumption that there will be a first patch for which Vicky returns a verdict different from ‘Yes’. As has already been pointed out in section 6.4.2, the assumption follows by classical logic from (CLEAR) and (MAX). The objection would then consist in pointing out that, given that ‘red’ is vague, classical logic cannot be legitimately assumed to hold in this context without further argument. Indeed, if that assumption cannot be made and, more generally, it cannot be assumed that Vicky adopts different attitudes with respect to patches which are at a close enough distance from one another (even if the distance should be larger than 1), it is hard to see how something like the problem we have been investigating could arise in the first place.

Appealing as it may seem, this train of thought misses one of the absolutely crucial features of the forced march. For the predicate that our

presentation of the forced march assumed to be subject to classical logic is not the vague ‘ ξ is red’, but simply ‘Vicky returns a ‘Yes’-verdict for ξ ’, and there is no reason to believe *this* predicate to be vague *in the case under consideration*. Of course, generally speaking, the predicate *is* vague, as we can conceive of a soritical series such that Vicky indisputably returns a ‘Yes’-verdict for the first object of the series and indisputably does not return a ‘Yes’-verdict for the last object of the series, with each two adjacent elements of the series x_i and x_{i+1} being such that Vicky’s act with respect to x_i differs from Vicky’s act with respect to x_{i+1} only by a nanometrical difference of the location of one single atom of Vicky’s body. For any such series, we are now not able to provide a warranted identification of a sharp boundary between the objects for which Vicky returns a ‘Yes’-verdict and the objects for which she does not; for any such series, we are now not able to decide for every object whether Vicky returns a ‘Yes’-verdict for it or not. For any such series, the predicate indeed appears to be vague over it and, correspondingly, the claim that there is a sharp boundary appears dubious (indeed false according to the theory defended in this essay).

However, it is essential to observe that no such vagueness needs to be present in the series over which Vicky is forced to march. For we can strengthen (HAPPY) with the following additional condition:

(SHARP) For every patch x , either it is indisputably the case that Vicky returns a ‘Yes’-verdict for x or it is indisputably the case that Vicky does not return a ‘Yes’-verdict for x .

There would then presumably be no obstacle to concluding that there will be a first patch for which Vicky returns a verdict different from ‘Yes’, and the presentation of the problem could proceed as before. We would only need to note that the additional imposition of (SHARP) seems unobjectionable: it would seem hardly satisfactory if one’s intellectual integrity could only be preserved by burbling.

Matters are however slightly more complex here. For (SHARP) amounts in effect to enforcing classical logic on the reactions to what are vague states-of-affairs. Any reason to think that classical logic does not apply to such states is thus transformed by the imposition of (SHARP) into a reason to think that the reactions in question are not both infallible and omniscient (in a word, that they are not “*transparent*”).²³ To see this, we consider a suitable soritical series and assume—plausibly enough—that at least one of the specific reasons to think that classical logic does not apply to vague states-of-affairs is that we do not want to conclude to the existence of a sharp boundary between positive and negative cases simply from the existence of both positive and negative cases. But if there are both positive and negative cases, there will be (by omniscience) a ‘Yes’-verdict for a positive case and there will not be (by contraposition on infallibility) a ‘Yes’-verdict for a negative case. By classical logic, there is a last case x for which there will be a ‘Yes’-verdict, and so (by infallibility) x is a positive case while (by contraposition on omniscience) the immediate successor of x is a negative case. Any reason for rejecting this conclusion is thus a reason for rejecting the joint assumption of infallibility and omniscience (and their contrapositives) for the reactions in question. (SHARP) cannot thus always hold for a transparent subject: its imposition is sometimes bound to generate either *error* (i.e. failure of infallibility) or *ignorance* (i.e. failure of omniscience). It follows that, over suitable soritical series, transparent subjects fail to meet (SHARP).²⁴

For transparent subjects, failure of (SHARP) is the reflection of a *positive* epistemic status (such as the one afforded by transparency) and so justified by it: the subject fails to meet (SHARP) simply because she is doing *so*

²³A subject’s reaction of \mathcal{A} ing to a state-of-affairs that P is *infallible* iff, if the subject \mathcal{A} s that P , then P ; it is *omniscient* iff, if P , then the subject \mathcal{A} s that P . I presuppose a fully analogous definition for *ranges* of states-of-affairs rather than *single* states-of-affairs.

²⁴Hawthorne [2005] explores a closely related dialectic, reaching similar conclusions. I’m especially indebted to Dan López de Sa and Sebastiano Moruzzi for challenging conversations on this topic.

well epistemically as to be transparent. One of the highest epistemic states, transparency, enjoins burbling. Such is vagueness. However, for at least some soritical series, there is no reason to think that Vicky will enjoy transparency with respect to them (indeed, no reason to think that she will enjoy either infallibility or omniscience), and so no reason to think that a similar motivation for failure of (SHARP) is available for her: her burbling cannot be interpreted as the mark of a higher epistemic state, and will thus presumably be subject to the default criticisms that an inquirer's burbling rightly elicits. In particular, this will presumably jeopardize Vicky's intellectual integrity, and so we are entitled to assume that, if this is to be preserved, (SHARP) must be satisfied.

There is another telling reason for thinking that appeal to vagueness for motivating a rejection of (SHARP) is misplaced in this dialectic (cf Sorensen [2001], pp. 40–56, whose general treatment of the forced march is quite congenial to the solution I will propose in section 6.5.1). That appeal requires a relevant expression (in our case, 'red') to be vague—however, a structurally identical problem could be created for Vicky by using only precise expressions. Consider a series of 1,000 pairwise indiscriminable trees, going from 25 ft tall to 15 ft tall. The trees are pairwise indiscriminable in the sense that, given any pair of adjacent trees, no normal human subject could detect, by unaided vision, any difference in height between the two trees when considering them in isolation from all other trees (either sequentially or at the same time). Consider a normal human subject, again Vicky, who, under the best visual conditions, is forced to give her best judgement sequentially on all the trees, starting from tree #1. Vicky may know about all this situation, and so know that at least some of the indiscriminable trees x and y could differ from one another at least in the sense of there being another tree z in the series such that z is discriminable from x but not from y (although the existence of such a tree is not guaranteed in the series). Still, Vicky cannot detect, by unaided vision, any difference in height between any two adjacent trees when considering them in isolation from all other trees. The question

that she will be asked at each step is ‘Is this tree at least 20 ft tall?’. Again, Vicky’s aim is to give a warranted answer at each step of this new forced march.

To this effect, she has available any answer whatsoever: ‘Yes’, ‘No’, ‘I can’t say’, ‘It’s borderline’, ‘Yes and no’ etc. We can then impose a constraint analogous to (HAPPY) for the preservation of Vicky’s intellectual integrity and go through a dialectic analogous to the one we have just been studying. The final move of rejecting the analogue of (SHARP) on the grounds of the vagueness of a relevant expression will however not be available, since the predicate ‘at least 20 ft tall’ is (for all intents and purposes) precise. *Inexact knowledge* of a certain property of being F (knowledge afforded by cognitive processes which unimprovably allow to recognize some case of being F without allowing to recognize every case of being F)²⁵ is thus a sufficient condition for the generation of the problem—vagueness is not a necessary condition for it. A fairly widespread view in the contemporary debate on vagueness has it that the forced march represents a paradox *of vagueness independent* from the sorites paradox, in the sense that, even though it equally depends on the vagueness of the relevant expressions, its paradoxicality is not entailed by (nor entails) the paradoxicality of the sorites reasoning, and that a theory of vagueness may well fare differently when applied to one paradox and when applied to the other one (in such a way that the forced march becomes a new crucial test that a theory of vagueness has to pass). That view, I hope to have shown, is wrong. The forced march is not in any interesting sense a paradox *of vagueness*.²⁶

²⁵Equivalently (under some assumptions), inexact knowledge of an object x is knowledge afforded by cognitive processes which unimprovably allow to recognize some determinable of x without allowing to recognize every determinable of x .

²⁶Granted, one might still be exercising vague concepts in the process of evaluating which judgement one should give with respect to the precise content that a certain tree x is at least 20 ft tall. One might for example bring to bear considerations concerning whether x is *close enough* in height to something which one is independently warranted in believing to be at least 20 ft tall. It is very unclear however whether even in such a

This should actually have already been apparent from our presentation of the original problem using the predicate ‘red’, since that problem simply arose from the epistemic indiscriminability for Vicky of the patches’ shades, not from any ignorance on Vicky’s part peculiarly attaching to the predicate ‘red’, and since epistemic indiscriminability of close enough values (such as specific shades and heights) along a certain dimension of comparison need not involve any vague categorization of the values of that dimension (such as categorization into colours or rough sizes). Of course, a different presentation of the problem could stipulate away the epistemic indiscriminability of close enough values. The series in question could e.g. be the series of the first 1,000 positive integers presented under their canonical mode of presentation. Every pair of adjacent numbers would then be discriminable in the sense that a normal human subject could detect, by unaided reflection, a difference in cardinality between the two numbers even when considering them in isolation from all other numbers (either sequentially or at the same time). We could then consider a normal human subject, again Vicky, who, under the best conditions for arithmetical reflection, is forced to give her best judgement sequentially on all the numbers, starting from number 1. The question that she will be asked at each step is ‘Is this number small?’ (we assume a fixed context where 1 counts as an indisputable positive case of ‘small’ and 1,000 as an indisputably negative case of it). Again, Vicky’s aim is to give a warranted answer at each step of this new forced march.

case vagueness would still be vindicated to be the source of the paradox: after all, the fact that one is also using numerical concepts does not lend any support to the claim that they are the source of the paradox. In any event, even if one might so use vague concepts, it seems plausible that one need not to and might just as well only rely on one’s perception and one’s understanding of what it is for something to be at least 20 ft tall. A being with the same perceptual limitations as ours who possesses the concept of being at least 20 ft tall but no relevant vague concept would be no better off than Vicky in the precise forced march. I don’t see any clear incoherence in the assumption of the possibility of such a being. Thanks to Crispin Wright for raising this important issue, which certainly deserves further exploration.

To this effect, she has available any answer whatsoever: ‘Yes’, ‘No’, ‘I can’t say’, ‘It’s borderline’, ‘Yes and no’ etc. We can then impose a constraint analogous to (HAPPY) for the preservation of Vicky’s intellectual integrity and go through a dialectic analogous to the one we have just been studying. But the dialectic would now seem to be entirely driven by the vagueness of ‘small’, for what else could justify our intuitive judgements of unwarrantability concerning the distinctions Vicky would now be forced to make? On the strength of these considerations, it could then be conjectured that vagueness is a necessary condition for the generation of the problem at least when other natural supplementary assumptions are made (such as the epistemic discriminability of close enough values along the relevant dimension of comparison). Plausible as this conjecture may seem, it too is bound to fail. For even in the case under discussion, a forced march can be constructed by using a precise predicate such as ‘ ξ bricks would break a camel’s back’. Even if Vicky can now discriminate between any two adjacent numbers, and can classify differently 1 and 1,000 under the categorization induced by ‘ ξ bricks would break a camel’s back’, she is still in no position warrantably to adopt any pair of differential attitudes towards any two adjacent numbers, in spite of the fact that, in this case, there is no reason to doubt that the categorization so induced does give rise to a sharp boundary between positive and negative cases.

Of course, *there is* a substantial difference between the precise and the vague case: whereas in the precise forced march there is no problem in thinking that the adoption of differential attitudes towards adjacent objects, although unwarranted, may well be lucky enough as to reflect correctly a sharp distinction in reality,²⁷ in the vague forced march there is a strong (and, at least according to the theory defended in this essay, correct) intuition that

²⁷At least considering attitudes whose contents remain at the level of positive or negative predications involving the original precise predicate—vagueness may of course come in as soon as “reflective” attitudes are adopted which concern e.g. the subject’s epistemic state (such as the attitude expressed by an utterance of ‘I don’t know’).

any such adoption is not only unwarranted, but *radically misconceived*—it requires a sharp distinction in reality where there could not possibly be one. It was thus essential to show that a deep problem for (E) arises even in the absence of this further condition.

Moreover, it is certainly very plausible that the problem is made even more acute by the presence of this further condition: vague forced marches impose violations of (E) which, because of their apparent radical misconception, are very plausibly even graver than those imposed by precise forced marches. Note that this circumstance lends however no support to the view that the forced march represents a self-standing paradox of vagueness: for the only new feature that vagueness brings in—the apparent radical misconception represented by the adoption of differential attitudes towards adjacent objects—is wholly grounded in the apparent radical misconception represented by the negation of the major premise of the sorites paradox. The additional paradoxicality of vague forced marches is thus simply a consequence of the paradoxicality of the corresponding sorites reasoning—we have not yet found a vagueness-generated paradoxicality independent from that exhibited by the sorites paradox.

6.4.6 Naive Forced March

This is of course not to deny that, given the eminent plausibility of (E), the forced march *is* a *paradox* (a case where despite the apparent validity of the argument, the apparently true premises do not appear rationally to support the conclusion). However, given the previous considerations, we should neither expect nor require from a theory *of vagueness* that it offers a solution (or, more weakly, a diagnosis) to the forced-march paradox flowing from its solution (or diagnosis) to the sorites paradox. From this perspective, it should thus be neither a surprise nor a discomfort to find out that the naive theory of vagueness itself is just as affected by the paradox as any other theory.

Prima facie, the naive theory, once supported by a tolerant logic (see chapter 4) and enriched with the conceptual resources required by the adoption of such a logic (see chapter 5) would seem fit to give a satisfactory solution not only to the sorites paradox, but also to the forced-march paradox. Let us suppose for simplicity's sake that the margin of tolerance for 'red' in our series is exactly 1 (under the natural measure function). Then Vicky, now a believer in the naive theory, may well accept 'Patch #400 is red' only because she accepts 'Patch #399 is red' for non-soritical reasons,²⁸ accepts the tolerance conditional 'If patch #399 is red, so is patch #400' for non-soritical reasons and recognizes the validity of *modus ponens*. Her reasons for accepting 'Patch #400 is red' are thus soritical, and so she is not committed to accepting 'Patch #401 is red', even though she accepts the tolerance conditional 'If patch #400 is red, so is patch #401' for non-soritical reasons and recognizes the validity of *modus ponens* (see section 5.4). Not being committed to accepting 'Patch #401 is red' by her antecedent beliefs, she may well adopt a different attitude towards patch #401 (even if maybe not that of accepting 'Patch #401 is not red', for example if she accepts (NEC) of section 4.3.3), while still accepting 'If patch #400 is red, so is patch #401', and so while still maintaining that patch #400 and patch #401 do not at all differ in their colour (if one is red, so is the other). Switch at a certain patch would then seem compatible with recognition that no relevant difference in colour exists between that patch and its predecessor. Has the naive theory managed a miraculous squaring of the circle?

Arguably no. Under the previous rules, Vicky must believe that, when she returns a different verdict for patch #401, she is still complying with (HAPPY) and all of its sub-conditions. The main line of argument of section 6.4.3 could then once again be applied to obtain the conclusion that she

²⁸For the purposes of the discussion in this chapter, we can take a subject *s* to accept φ for *soritical* reasons iff *s* accepts φ because *s* accepts a tolerance conditional whose consequent φ is and also accepts its antecedent (see section 5.4.3). Note that, contrary to the usage of chapter 5, here I will use 'reason' and its like only to refer to something which in effect provides some positive epistemic support for a certain belief.

both believes that [she is warranted in judging patch #400 to be red] and believes that [she is not warranted in judging patch #401 to be red], *and so* that she believes that [she is warranted in judging patch #400 to be red and not warranted in judging patch #401 to be red], a belief for which we have already argued that Vicky has no warrant.

To this it may be replied that, in the novel logical context of tolerant logics, the last (adjunction) step of the previous argument presupposes that Vicky's reasons both for believing that she is warranted in judging patch #400 to be red and for believing that she is not warranted in judging patch #401 to be red are non-soritical, for otherwise adjunction may not be guaranteed to generate a valid argument from the original premises to the target conclusion (that is, adjunction may not have the full 1-left-transitivity property, see section 4.2.3). The naive strategy under consideration may then retort that this presupposition is groundless: Vicky may well believe that she is warranted in judging patch #400 to be red only because she both believes (for non-soritical reasons) that she is warranted in judging patch #399 to be red and believes (for non-soritical reasons) that, if she is warranted in judging patch #399 to be red, she is also warranted in judging patch #400 to be red. Analogously, Vicky may well believe that she is not warranted in judging patch #401 to be red only because she both believes (for non-soritical reasons) that she is not warranted in judging patch #402 to be red and believes (for non-soritical reasons) that, if she is not warranted in judging patch #402 to be red, she is also not warranted in judging patch #401 to be red. In this case, letting \mathbf{T}^* be the naive theorist's favoured tolerant logic, the adjunction step would be an application of (T^l) to $\varphi_0 \supset \varphi_1, \varphi_0 \vdash_{\mathbf{T}^*} \varphi_1$ and $\varphi_2 \supset \varphi_3, \varphi_2 \vdash_{\mathbf{T}^*} \varphi_3$ to yield $\varphi_0 \supset \varphi_1, \varphi_0, \varphi_2 \supset \varphi_3, \varphi_2 \vdash_{\mathbf{T}^*} \varphi_1 \wedge \varphi_3$, application which may not be legitimate in \mathbf{T}^* .

Let us leave aside the fact that it is actually possible that a suitable tolerant logic declares that the adjunction step in question is guaranteed to generate a valid argument from the original premises to the target conclusion (in fact, even though none of the main systems of chapter 4 does it, this

could easily be achieved by imposing more structure on the set of tolerated but not designated values). And let us also leave aside the fact that even if the pattern of reasoning in question were formally invalid in every tolerant logic, it would still have to be shown that its current application, despite its apparent plausibility, is not materially valid. The main problem with the present reply is rather that, contrary to what it presupposes, Vicky does have non-soritical reasons both for believing that she is warranted in judging patch #400 to be red and for believing that she is not warranted in judging patch #401 to be red: given that she in fact judges patch #400 to be red and does not judge patch #401 to be red, the main line of argument of section 6.4.3 is available to give her non-soritical reasons for both beliefs.

These are however once again relatively minor niceties. For, as in the case of the objection to collection of warrant over conjunction of section 6.4.2, the simple and yet damning point is that Vicky's pair of attitudes—belief (for soritical reasons) that she is warranted in judging patch #400 to be red, belief (for soritical reasons) that she is not warranted in judging patch #401 to be red—draws a sharp distinction between patch #400 and patch #401 (with regard to their being warrantably judgeable by Vicky to be red) for which Vicky has no warrant: for all Vicky can tell, an analogous pattern of attitudes placing the sharp boundary in question rather between, say, patch #401 and patch #402 would have been just as good (or just as bad).

Having thus disposed of the naive theory of vagueness *as a viable strategy to solve the forced-march paradox*, it is important to see how a forced-march-related objection against the theory *in itself* fails. The objection is to the effect that a believer in the naive theory will believe (for non-soritical reasons) that patch #1,000 is red as soon as she believes for non-soritical reasons that patch #1 is red, and goes as follows:

ANTI-NAIVE Consider a believer in the naive theory of vagueness, Monsieur Naïf, and consider the predicate ‘ ξ is such that Naïf has non-soritical reasons to believe that ξ is red’ (henceforth, for short, ‘ ξ is

Naïf-non-soritical’). This predicate is just as vague as ‘ ξ is red’. Assuming that Naïf is minimally competent, there will be by assumption at least one patch (patch #1) believed by him for non-soritical reasons to be red, and so, very plausibly, at least one patch believed by him for non-soritical reasons to be Naïf-non-soritical. This last inference is certainly something the naive theorist should be willing to concede as correct for patch #1 and for quite a few other patches after it. So take any such patch # i . Naïf believes for non-soritical reasons that patch # i is Naïf-non-soritical. Believing in the naive theory of vagueness, Naïf also believes for non-soritical reasons, given the vagueness of ‘ ξ is Naïf-non-soritical’, that, if patch # i is Naïf-non-soritical, so is patch # $i + 1$. Therefore, assuming a *modicum* of rationality, Naïf believes for *soritical* reasons that patch # $i + 1$ is Naïf-non-soritical.

Now, at least in a highly idealized context where the only information available is that provided by reflection on the perceptually accessible colour of the patches, we may well take a reason to be non-soritical iff it is *canonical* (that is, yielded by simple reflection on the colour of the relevant patch considered in isolation from the colour of any other patch). Moreover, given the idealization in question, it would seem that the *only* way for Naïf to gain the information that he has canonical reasons to believe that patch # $i + 1$ is red is to gain access to one of these reasons, from which it certainly follows, assuming again a *modicum* of rationality, that Naïf believes for non-soritical reasons that patch # $i + 1$ is Naïf-non-soritical. It thus follows that Naïf can only believe for reasons that patch # $i + 1$ is Naïf-non-soritical if he believes (also) for *non-soritical* reasons that patch # $i + 1$ is Naïf-non-soritical. We have thus proved that, if Naïf believes for non-soritical reasons that patch # i is Naïf-non-soritical, he also believes for non-soritical reasons that patch # $i + 1$ is Naïf-non-soritical. From this, using a classical metalanguage (which is legitimate, since ‘Naïf believes for non-soritical reasons that ξ is Naïf-non-soritical’ can be taken to be precise

if Naïf is subject to an analogue of (SHARP) for ‘Naïf-non-soritical’, the soritical/non-soritical distinction is precise and the relevant instances of the basing relation²⁹ are also precise), the desired result follows from familiar reasoning.

It should by now be clear where **ANTI-NAIVE** goes astray. It goes astray in assuming that the only way for Naïf to gain the information that he has canonical reasons to believe that patch $\#i + 1$ is red is to gain access to one of these reasons. Since Naïf believes in the naive theory of vagueness, this amounts to overlooking the crucial possibility that Naïf gains that information by inferring it from the information that he has canonical reasons to believe that patch $\#i$ is red in conjunction with the relevant tolerance conditional—possibility which was even mentioned in the previous leg of the argument!

Despite this fallacy, **ANTI-NAIVE** retains great interest in showing what is indeed a remarkable consequence of believing in the naive theory of vagueness. As soon as competent and rational Naïf renounces his transparency by subjecting himself to the relevant analogue of (SHARP) (and the other assumptions on precision mentioned at the end of **ANTI-NAIVE** are in place), he will find himself committed to believing (for soritical reasons), for some i , that patch $\#i$ is Naïf-non-soritical without being able to access himself any of the relevant non-soritical reasons (and so without being able to transform his soritical reasons for that belief into non-soritical ones), even assuming that what appear to be the best possible conditions for a subject to access these reasons hold. The problematicity of the situation is made even more acute when we consider predicates for *phenomenal* properties such as ‘ ξ_0 looks red to ξ_1 at ξ_2 ’: at some time t , Naïf will find himself committed to believing (for soritical reasons), for some i , that patch $\#i$ is such that Naïf has non-soritical reasons to believe that patch $\#i$ looks red to him at t without being able to access himself any of the relevant non-soritical reasons.

²⁹The *basing relation* is the relation which holds between a belief b of a subject s and s ’s reasons Rs iff s holds b for Rs (see Korcz [1997] for a recent survey).

But what can prevent Naïf from accessing these reasons relating to patch $\#i$'s look given that the conditions appear to be the best possible ones for doing so (patch $\#i$ is fully open to view, Naïf's perceptual system is working optimally, Naïf is in the clearest possible state of mind for reflecting about patch $\#i$'s look etc.)?

The presentation of the problem admits of slight variations. Define ' ξ is Naïf-non-soritical*' as ' ξ is such that Naïf believes for non-soritical reasons that ξ is red' (' ξ is Naïf-non-soritical*' differs from ' ξ is Naïf-non-soritical' in that, given Naïf's subjection to (SHARP) and the other assumptions on precision mentioned at the end of **ANTI-NAIVE**, the former is precise whereas the latter is vague). Given the precision of this predicate, there will be a last Naïf-non-soritical* patch $\#i$. Believing in the naive theory, Naïf also believes for non-soritical reasons, given the vagueness of 'red', that, if patch $\#i$ is red, so is patch $\#i + 1$. Therefore, Naïf believes for soritical reasons that patch $\#i + 1$ is red. Moreover, since, by assumption, patch $\#i$ is the last Naïf-non-soritical* patch, Naïf does not believe for non-soritical reasons that patch $\#i + 1$ is red, even assuming that what appear to be the best possible conditions for a subject to access non-soritical reasons hold. And, again, the problematicity of the situation is made even more acute when we consider predicates such as ' ξ_0 looks red to ξ_1 at ξ_2 '.

There is no denying that such situations are uncomfortable for the subject who finds herself in them, but they should by no means be seen as paradoxical, or as presenting a challenge to the naive theory. They crucially arise because Naïf is subjected to (SHARP) itself or to some of its analogues, and we have already seen in section 6.4.5 that such a subjection is incompatible with Naïf's transparency with respect to the relevant states-of-affairs. Since, on both versions of the problem, this consists in Naïf's being blind to the existence of certain non-soritical reasons, we can assume that the failure of transparency induced by subjection to (SHARP) or its analogues is actually a failure of omniscience rather than infallibility (see fn 23).

Under this assumption, Naïf's blindness to the existence of the relevant

non-soritical reasons can neatly be explained as follows. Naïf's subjection to an analogue of (SHARP) for 'Naïf-non-soritical' is incompatible with Naïf's transparency (i.e. omniscience) with respect to there being non-soritical reasons to believe that a certain patch is red. This, in conjunction with the plausible assumption that, in Naïf's highly idealized context, any non-soritical reason to believe that a certain patch is Naïf-non-soritical would be a non-soritical reason to believe that a certain patch is red, explains why, in the first version of the problem, Naïf cannot access any non-soritical reason to believe that patch $\#i$ is Naïf-non-soritical. Moreover, Naïf's subjection to (SHARP) itself is incompatible with Naïf's transparency (i.e. omniscience) with respect to a certain patch's being red. This, in conjunction with the plausible assumption that, in Naïf's highly idealized context, if Naïf has access to a certain non-soritical reason, he also has access to every soritical reason generated by it (so that failure of Naïf's omniscience with respect to a certain patch's being red in effect amounts to failure of Naïf's omniscience with respect to there being non-soritical reasons to believe that a certain—possibly antecedent—patch is red), explains why, in the second version of the problem, Naïf cannot access any non-soritical reason to believe that patch $\#i + 1$ is red (to stress, because otherwise Naïf would be omniscient with respect to patch $\#i + 2$'s being red, which he may well not be). Given this forced failure of transparency (i.e. omniscience), it is thus just what one should expect that, for some patches x, y , Naïf is blind to the existence of non-soritical reasons to believe that x is Naïf-non-soritical, and blind to the existence of non-soritical reasons to believe that y is red.

What, among other things, the foregoing shows is in effect that, under the current assumptions concerning Naïf, at least some proposition saying that a certain patch is red is non-deductively inaccessible for Naïf at the time of the forced march (see section 5.4.3). Using equivalent assumptions, we could have derived the same conclusion from the more general result of section 5.4.3 concerning the connection between non-deductively inaccessible propositions and at least intermediate applications of non-transitivity. What is crucial

to stress in closing this section is that none of these findings lends support to the objection that a naive theorist like Naïf is committed to holding that an object x can be, say, red without there being any canonical reasons to believe that it is so (because, so the objection goes, even though we cannot establish *by looking* that x is red, we can still do so, *qua* naive theorists, *by inferring* that x is red from a relevant tolerance conditional together with its antecedent). This is so because, for all that has been shown in our discussion of **ANTI-NAIVE**, a naive theorist can still hold that [something is red iff there are canonical reasons to believe that it is so] (from which, together with the relevant tolerance principle for ‘red’, it plausibly follows that, if there are canonical reasons to believe that patch $\#i$ is red, there are also canonical reasons to believe that patch $\#i + 1$ is red). This does not contradict any of the previous findings, as long as we keep in mind that the imposition of (SHARP) and its analogues on a certain subject s forces a gap between the existence of canonical reasons to believe a certain proposition and these reasons’ accessibility to s .³⁰

6.5 The Penumbra

6.5.1 Falling Asleep with an Unhappy Face

Given the failure of even the most promising strategies, we are entitled to conclude that forced-marched Vicky will not be able to return a warranted answer for every patch in the series. (E) fails—Vicky cannot possibly preserve her intellectual integrity while judging every patch in the series. Such is the forced march. No matter how repugnant this conclusion used to appear to us before a detailed examination of the forced march, it *is* the conclusion which is mandated at the end of such examination, and I will henceforth assume it.

More specifically, let us distinguish between two (related) paradoxes. In

³⁰Thanks to Crispin Wright for illuminating discussions of the cluster of issues surrounding **ANTI-NAIVE**.

one sense, our paradox was that (E) is apparently true, and nevertheless consideration of the forced march under apparently true assumptions would appear to entail that it is false. The straightforward—even though maybe disappointing—solution to that paradox is that, despite its apparent truth, (E) is after all false (let us call this the ‘*forced-answer paradox*’). But there is another, arguably deeper sense in which our paradox was that it is apparently true that one can always preserve one’s intellectual integrity, and nevertheless consideration of the forced march under apparently true assumptions would appear to entail that this is false (let us call this the ‘*forced-fault paradox*’). Let us also distinguish between two kinds of solutions to the forced-fault paradox: on the one hand, “*happy-face solutions*”, which try to solve the paradox by respecting (E), and so by devising an overall pattern of answers which would preserve Vicky’s intellectual integrity while allowing for her to judge every patch in the series; on the other hand, “*unhappy-face solutions*”, which try to solve the paradox by rejecting (E), and so by disallowing Vicky to judge every patch in the series.³¹ Given our assumption that (E) is false, we can thus only hope for an unhappy-face solution to the forced-fault paradox. What should one do when one is forced-march? After all, given that (E) is false, there is no strategy one can undertake which would allow one to preserve one’s intellectual integrity while fully engaging in the forced march. So, assuming that one’s intellectual integrity is unnegotiable, what should one do?

That “there is no strategy one can undertake which would allow one to preserve one’s intellectual integrity while fully engaging in the forced march” just means that there is nothing one can do or try to do which would effectively preserve one’s intellectual integrity while fully engaging in the forced march. Any action *of this kind* would be in vain. This might suggest that

³¹I’m thus going to use the expressions ‘happy-face/unhappy-face solution’ in a different semi-technical sense from the semi-technical sense adopted by Stephen Schiffer in his influential Schiffer [1998]; Schiffer [2000]; Schiffer [2003], pp. 178–237. I’ve decided to stick to these labels because of their compelling vividness.

some action *of a different kind* would be successful. Indeed, it is often intimated as an alternative, more effective strategy to the ones we have been reviewing that at some point in the forced march one should simply *opt out of the game*, refusing to keep on answering the malign questions of the interrogator. At some point in the forced march, one should simply shrug one's shoulders. This strategy is typically offered in an unhappy-face spirit, as the only thing which is left to do for Vicky given that all the happy-face strategies—which try to meet head-on the interrogator's questions—are deemed to be bound to fail. Indeed, the arguments in section 6.4 would then break down at least in their letter, since they assume that there will be a first patch for which Vicky returns a verdict different from 'Yes' immediately succeeding the last patch for which Vicky returns 'Yes' as a verdict, and this assumption is (classically) justified in the setting of the forced march only if Vicky does return a verdict of some kind or other for all the patches in the series.

However, the arguments in section 6.4 would still remain unscathed in their spirit. For once the deep reasons are appreciated as to why the happy-face strategies are bound to fail (reasons which I have tried to unearth in section 6.4), it should be clear that the unhappy-face strategy of shrugging one's shoulders is no better off than happy-face strategies are. For Vicky can be subjected to an analogue of (SHARP) for shrugging her shoulders at patch x (rather than for returning a 'Yes'-verdict for x), and so there will be a first patch $\#i$ at which Vicky shrugs her shoulders. But then the question arises as to why Vicky starts shrugging her shoulders at patch $\#i$ rather than at, say, patch $\#i - 1$ or patch $\#i + 1$, and a dialectic analogous to the one developed in section 6.4 will occur, with the same negative conclusion. The dialectic will occur because, even if she decides to shrug her shoulders at the question as to whether patch $\#i$ is red, Vicky can hardly shrug her shoulders at the question as to why start shrugging one's shoulders exactly at patch $\#i$ (given that she does so!).³²

³²True, actions can quite generally be described at such a level of specificity that it

Of course, one could decide to disengage from the forced march in the radical sense of not taking part in it in the first place: one would not form any judgement at all about any patch in the series, not even about patch #1. But in what sense would such a draconian measure manage to preserve one's intellectual integrity? One would not even recognize an indisputable positive case of red as such, but, as we have already noted in section 6.4.4, this failure of recognition in the face of overwhelming positive or negative warrant is a parody of inquiry. Setting thus aside this last suicidal option, our paradox is now revealed to be even more dramatic: if one cannot even so much as opt out of the forced march while preserving one's intellectual integrity, what should one do?

One shouldn't *do* anything. The previous arguments show that, if one really *considers* all the questions which naturally arise in the course of a forced march (or *self-consciously dismisses* some of them), in one way or another one is doomed to lose one's intellectual integrity. *Any* action would be in vain. Thus, one can only hope to *forget* about at least some of these

would make little sense to demand for an action to be justified that the agent be able to justify, at that level of specificity, her acting in a certain very specific way rather than in a different but barely distinguishable one. For example, it would make little sense to demand for your leaving the room to be justified that you be able to justify, at the level of nanoseconds, your leaving at a certain very specific time t (as, barring irrelevant vagueness, you unavoidably will) rather than a nanosecond earlier or later than t . And crucially, in the present context, it becomes irrelevant that mental actions such as judgements seem to represent a counterexample to the unrestricted validity of the previous claim (presumably because a *mental* action—as opposed to a *material* one—can remain unspecific in the relevant respects by directing itself towards a suitably unspecific content), since shrugging one's shoulders is a material rather than a mental action (I'm here helping myself to what I hope is an intuitive and non-committal understanding of the mental/material distinction). Sharpening the way the point is put in the text, I reply that, as long as shrugging her shoulders is a self-conscious action on Vicky's part, it will involve the formation of judgements concerning whether a certain patch is such that it would be reasonable to shrug her shoulders at it. At least with respect to these judgements (and under the aegis of (COH)), a dialectic analogous to the one developed in section 6.4 will occur, with the same negative conclusion. Thanks to Stephen Read for raising this worry.

questions. This could happen in several ways: by failing to notice that they do arise, by getting distracted into other activities, by falling asleep etc. Crucially, these are not *actions* one *undertakes*: any such self-conscious action to the effect of blocking the consideration of the relevant questions for at least some patch would become involved in the dialectic just exposed. These and other species of forgetting about a question are rather *events* that *happen* to one, and all one can do is to hope that they will happen.

It is because they are such blind happenings that a *human* inquirer cannot be criticized on *intellectual* grounds for undergoing them, even though she may be criticized for that on *practical* grounds (for example, for not taking a pill which would have kept her awake). Such blind happenings are thus compatible with the preservation of the inquirer's intellectual integrity. One cannot even make preparations for them to happen, as these preparations would need to involve instructions (like 'Take a pill that will make you fall asleep after patch #350') which, in order to avoid the previous draconian alternative, would have to make sure that one does not forget to judge at least in some cases—but the by now all too familiar question would then once again arise as to how the line can be drawn in a warranted way... Uncontrolled forgetting is the only safe way out of the forced march—or so I must conclude with an unhappy face.

6.5.2 Doxastic Paradox

It must be recognized that our proposed solution to the forced-fault paradox may well have appeared repugnant to us before a detailed examination of the forced march. Nevertheless, it *is* the solution which is mandated at the end of such examination—all better-looking alternatives have revealed flawed. The argument for our solution has in effect been an unabashed argument by elimination, in which we gleefully shot down alternative happy-face and unhappy-face solutions. I think however that, in the presence of the undeniable initial repugnancy, we can provide additional support for our so-

lution by showing that a similar solution is the correct one in another area independent from the forced march. Again, the area we will be looking at is the same as the one we explored from a different perspective in section 6.3.2 to refute (QS): the area of the paradoxes of self-reference, this time in particular the *doxastic paradoxes*.

Consider a modified version of the *Believer paradox* (re-discovered in modern times by Burge [1984]), with a sentence β_0 provably equivalent with ‘ β_0 is not believed by John’³³ (where ‘believe’ and its like are short for ‘believe at some time or other’, with suitable modifications under embeddings).

- (i) Suppose that John believes β_0 . Then, by double-negation introduction, it is not the case that it is not the case that β_0 is believed by John, and so, since by construction β_0 entails ‘It is not the case that β_0 is believed by John’, β_0 is not the case. Therefore, by structural contraction, John believes something that is not the case.
- (ii) Suppose that John does not believe β_0 . Then, since by construction ‘ β_0 is not believed by John’ entails β_0 , β_0 is the case. Therefore, by structural contraction, John does not believe something that is the case.

In either case, John would fail to keep track of reality, in particular of whether or not he believes β_0 . But, crucially (for our present purposes), the Believer paradox does not exhaust itself in this already uncomfortable situation. For reflect that the simple piece of reasoning in (i) is available to John. John has thus a simple conclusive reason not to believe β_0 . Given this reason, if he considers at all the question as to whether or not believe β_0 , John should thus come to the conclusion that he should not believe β_0 . But, if he considers at all the question as to whether or not believe β_0 , John can certainly also come to know that this latter piece of reasoning is available to

³³Again, to keep things simple, in making this point I use ‘believe’ and its like as predicates of sentences. Comments analogous to those made in fn 7 apply.

John (himself), and so, trusting John's (his) rationality, come to believe that John does not believe β_0 . Ouch!

This situation bears a striking resemblance to the forced-fault paradox. For here as well there seems to be a serious threat to John's intellectual integrity. This should already be apparent from the previous presentation, as the reasoning exploited there appears to be absolutely compelling (if John considers at all the question as to whether or not believe β_0), and yet commits John to believing something (β_0) that, in the course of the same reasoning, he has already committed himself not to believing. Since no one can both believe and fail to believe that something is the case, the preservation of John's intellectual integrity is impossible. A slight variation on the reasoning may make the analogy even more forceful. Either John believes β_0 or he doesn't. If he does and considers at all the question as to whether or not believe β_0 , given a *modicum* of self-knowledge he can run through (i) to undermine any warrant he might have had for so believing; if he doesn't and considers at all the question as to whether or not believe β_0 , given a *modicum* of self-knowledge he can run through (ii) to undermine any warrant he might have had for not so believing. Such is the Believer paradox.

The paradox cannot plausibly be blocked by revising the (non-trivial) logic used in it, because, in contrast to 'true' or 'known (by Thomas)', 'believed (by John)' is a predicate whose application *can be entirely decided on perfectly ordinary grounds*, even if the sentence under consideration is self-referential in the peculiar way paradoxical sentences are. Of course, it is not to be denied that there might be cases where the application of 'believed (by John)' is itself indeterminate (consider for example borderline cases for it), but such cases can be screened off by subjecting John to (SHARP) and its analogues (an analogous comment would apply to dialetheist proposals of solution). And there would seem to be no bar to the use of classical logic to a predicate whose application is so decidable.

Nor can the paradox be blocked by recourse to a hierarchy of belief predicates, so that, setting the *i*th-order to be the order of the belief predicate

occurring in β_0 , John should not believe^{*i*} β_0 but should believe^{*i+1*} β_0 . For the paradoxical reasoning is no less compelling when carried out using the belief^{*i*} predicate instead of the belief^{*i+1*} predicate: if he considers at all the question as to whether or not believe^{*i*} β_0 , John can certainly also come to know (and hence believe^{*i*}) that the relevant piece of reasoning is available to John (himself), and so, trusting John's (his) rationality, come to believe^{*i*} that John does not believe^{*i*} β_0 . If belief^{*i*} is such as to fail to make this reasoning compelling, it must fail to obey some norm which is deeply embedded in our pre-theoretical concept of belief, and so must all belief properties in the hierarchy, since, for every j , an analogous paradox can be devised for the belief^{*j*} predicate. Maybe the suggestion is really that, as far as belief^{*i*} goes, John should forget about the question as to whether or not believe^{*i*} β_0 —if so, I think the suggestion is on the right track, but the introduction of the hierarchy does not seem to achieve anything substantial additional to what the solution to be proposed achieves.

The paradox cannot even be blocked by caution about one's future rationality, so that John should not conclude from the availability to John (himself) of the simple piece of reasoning in (i) (and of the ensuing availability to John (himself) of the consequences of the former availability) that John (he himself) will always do what he has most reason to do, and so will never believe β_0 . For, whatever one might think of such a caution, an analogous paradox can be devised by using a sentence β_1 provably equivalent with ' β_1 is not now believed by John', and it would seem that John cannot but trust John's (his) present rationality.

As the presentation of the paradox already suggests, the only way out of it for John seems to be that of *not considering* the question as to whether or not believe β_0 . It is worth noting that the unhappy-face strategy of not considering in the sense opting out of the game presents itself here as well, and that it seems to suffer exactly from the same problem we discussed in section 6.5.1 for its application to the forced-fault paradox: if John self-consciously decides not to consider the question as to whether or not believe

β_0 , how can he avoid forming the belief that he does not believe β_0 ? A better unhappy-face solution seems to be that of forgetting about the question as to whether or not believe β_0 , in analogy to the forgetfulness solution that we defended with respect to the forced-fault paradox. Such an unhappy-face solution would break the spell of the apparent symmetry of the reasonings in (i) and (ii): forgetting about this question, John will certainly not believe β_0 (and so will not run afoul of the reasoning in (i)), but will also not be criticizable for not believing what is in fact the case (namely, β_0), for, after all, he has forgotten about that question (and so will not run afoul of the reasoning in (ii)).

6.5.3 The Source of Borderline Cases

From the standpoint of the traditional approach to the problem of vagueness adopted in this essay (see section 1.4), our favoured unhappy-face solution to the forced-fault paradox enjoys a considerable additional advantage. To recall, that approach takes as basic the phenomenon of sorites susceptibility, thereby standing in sharp contrast with the dominant approach which rather takes as basic the phenomenon of borderlineness, interpreting it as strong borderlineness. Yet, as detailed in section 1.3, there seems to be a solid intuition that, on each episode of consideration of a soritical series, some items cannot be decided (i.e. are such that the inquirer is neither in a position warrantably to believe that they are positive cases nor in a position warrantably to believe that they are negative cases)—in other words, that no soritical series, *even though finite in the classical sense*, can be *surveyed* all at once. Even though I reject the existence of strong borderline cases (see again section 1.3), I still accept the content of this borderlineness intuition, and I would regard it as a serious defect of the theory of vagueness defended in this essay if it were not able to account for it.

Fortunately, the forgetfulness solution to the forced-fault paradox provides the materials for accounting for the phenomenon of borderlineness

without appeal to any distinctive normative status enjoyed by some items in a soritical series (i.e. without appeal to strong borderline cases). For the solution requires that, on each episode of consideration of a soritical series, some items must be left forgotten, and so *a fortiori* cannot be decided. My proposal is then to identify the borderline cases in a soritical series S on an episode of consideration e of S with all and only those items in S that are forgotten on e in such a way as to make possible the forgetfulness solution to the forced-fault paradox.

Some *caveats*. Firstly, the proposal is neutral as to what to understand “episodes of consideration” more exactly to be: variations can be had both with respect to whom to count as subject of the consideration (a single subject, the participants in a context, a whole linguistic community) and with respect to what to count as an act of consideration (a sequence of explicit mental judgements, the pattern of beliefs at a certain time, the information potentially available at a certain time). With regard to the latter, note that the exact sense of ‘forgotten’ should also be taken to vary according to the particular choice made for the second parameter (not explicitly mentally judged, not believed, no information about). There is reason to think that many of the different notions generated by setting these (and maybe other) parameters to some of these (and maybe other) values will prove theoretically interesting—I myself am inclined to think that the notion generated by the middle values just mentioned will prove especially useful and close to some aspects of the pre-theoretical notion of a borderline case (see chapter 1, fn 3), and will henceforth assume it as the default notion.

Secondly, as should have emerged from the discussion of the forced march, by ‘ x is forgotten (on an episode of consideration e)’, ‘ x is not considered (on e)’, ‘ x lies beyond the horizon of one’s consideration (on e)’ etc. I simply mean that the question whether x is F or not (where ‘ F ’ is the predicate for which the series is soritical) is forgotten (on e). In particular, I do not mean that other properties of x are forgotten, let alone that x ’s very existence is forgotten.

Thirdly, note that the role of the clause ‘in such a way as to make possible the forgetfulness solution to the forced-fault paradox’ is that of narrowing down the borderline cases to the largest most central convex set of forgotten items, thereby approximating a salient feature of the pre-theoretical notion of a borderline case.

Fourthly, let me stress that my proposal is “*reconstructive*” rather than “*descriptive*”: I do not intend to say what the pre-theoretical notion of a borderline case (if there is a unique such thing) is (even though I have taken and shall take inspiration from a couple of features that are displayed in ordinary thought and talk), let alone what the notion of a strong borderline case (the notion used by the dominant approach) is. What I intend to do is rather, having identified on the basis of the phenomenon of borderlineness a theoretical role to play, to propose a notion which has naturally arisen in the context of the forced march and which seems fit to play that role. One would probably get even closer to the pre-theoretical notion of a borderline case by adding the clause that it must be a hard question to decide whether the item is a positive or a negative case (in a sense of ‘hard question’ which of course does not imply that the item is a strong borderline case). I’m inclined to adopt such a strengthening of the proposal, but nothing of importance here will hinge on this decision.

Fifthly, one can extend the proposal to definite cases, by saying that x is definitely F iff x is F and not borderline F (see section 3.3.1). This yields a well-behaved notion of a definite case—in particular, a notion which is factive and so allows the explanation of indefiniteness in terms of tolerance offered in section 1.4. I must leave it to another occasion to explore which properties of the property of being definite are induced by the proposal (for example, whether it iterates). Here, it will suffice to note that such an extension would arguably make mandatory what would already have seemed to be a desirable broadening of the original proposal: namely, to define the property of being borderline (and hence the property of being definite) not only for objects (with respect to properties), but also for propositions—in a familiar

terminology, to provide not only a *de re*, but also a *de dicto* reading for ‘borderline’-phrases (and hence for ‘definitely’-phrases).³⁴ Such a broadening can be achieved by introducing the notion of a *forgotten proposition* in the mould of the notion of a forgotten item. I leave the details to the reader.

Indeed, the extra strength of ‘forgotten’ is welcome, for it seems that the phenomenon of borderlineness goes as far as to warrant a protest against *any complete classification* of a soritical series. ‘Patch #1 is absolutely red; patch #2 is almost absolutely red; patch #3 is roughly absolutely red. . . patch #400 is almost neither definitely red nor definitely not red; patch #401 is just neither definitely red nor definitely not red; patch #402 is little more than just neither definitely red nor definitely not red. . . patch #600 is almost not red; patch #601 is just not red; patch #602 is little more than just not red. . . patch #998 is roughly absolutely not red; patch #999 is almost absolutely not red; patch #1,000 is absolutely not red’: it is natural to react to any such classification by saying ‘Look, you can’t do this—you must leave some borderline cases!’. In this and similar speeches, we can glimpse a feature of the pre-theoretical notion of a borderline case which has gone unnoticed for far too long: not something which enjoys a distinctive status, but something which, on the relevant episode of consideration, is not attributed any such status—forgotten, lying beyond the horizon of one’s present consideration.

Since I have argued in section 6.4.5 that the forced-fault paradox is not at all a paradox of vagueness, it should be expected that forgotten items will be present also on episodes of considerations of non-soritical, precise series like those considered in section 6.4.5. It should have already been expected that something very similar to (if not identical with) borderline cases arises for non-soritical series affected by some sort or other of indeterminacy—yet, it might be seem doubtful that borderline cases still arise even when all indeterminacy (be it vagueness-related or not) is absent. However, it

³⁴Ineliminable *de dicto* occurrences of ‘definitely’ have taken for example centre stage in chapter 3. I’m grateful to Neil Cooper for making me aware of the importance of this issue.

is certainly in the spirit of the theory of vagueness defended in this essay to allow that something very similar to borderline cases will arise also on episodes of considerations of precise series, for, on this theory, borderline cases do not enter at all into the definition of what vagueness is.

Be that as it may, I think that there *are* some important differences between forgotten items in a precise series and forgotten items in a soritical series which justify my prising them apart for theoretical purposes (note that my official definition of a borderline case applies only to *soritical* series and so picks out only the latter kind of forgotten items, but it is clear that without further substantiation this would be just a point of terminology). Forgotten items in a soritical series present an *opacity to inquiry* which is in some respect stronger and in some respect weaker than the opacity presented by forgotten items in a precise series.

On the one hand, the opacity is stronger in the sense that it cannot be eliminated by simply switching to already known, more powerful methods of inquiry, whereas it can be so eliminated in precise series (e.g. by recourse to ruler measurements in the case of the series of trees of section 6.4.5). Indeed, on most views on vagueness, the opacity is so strong in this respect that it cannot be eliminated at all, holding of necessity for every subject and method (see Sorensen [2001], pp. 21–39). In this specific respect, such views have been almost fully vindicated by the present analysis, since, as we have seen in section 6.4.5, even God himself, *if subjected to* (SHARP), would have to forget about some item (of course, given my rejection of the existence of strong borderline cases, I think that the necessary opacity in question is only, as it were, *de dicto* rather than *de re*: it is the necessity that, on every episode of consideration, some item or other remain forgotten, and not the necessity belonging to some particular item to remain forgotten on every episode of consideration).

On the other hand, the opacity is weaker in the sense that, keeping the method fixed, for each particular forgotten item in a soritical series, there is no reason to rule out that subjects belonging to another episode of considera-

tion do manage to decide the status of the item by that very same method (at least, this is so if borderline cases are not strong borderline cases), whereas, for some item in a precise series, there is every reason to rule out that any subject manages to decide the status of the item by that very same method.

In addition to the divergent answers to the question as to whether or not borderline cases enjoy a distinctive normative status, there are of course at least two other quite conspicuous differences between the proposed reconstruction of the notion of a borderline case and the interpretation given to it by the dominant approach. The first difference is that the proposed notion will be, in many cases of interest, *precise*, since, in many such cases, it will be a precise question which items are forgotten on which episode of consideration. Granted, sometimes this question will itself be vague (there can be a soritical series from forgotten items to unforgotten ones). Therefore, if “higher-order vagueness” is vagueness in one’s favoured notion of a borderline case (see section 1.3), there will still be on this view a marginal residue of higher-order vagueness. Nevertheless, in many cases of interest, it will be legitimate to assume that the borderline cases of a vague predicate have precise boundaries. In particular, in the context of the paradoxes of higher-order vagueness (see chapter 3), a new formidable argument needs to be mounted to the effect that there really is all the higher-order vagueness that the paradoxes require. The prospects for such an argument are at best bleak. Moreover, the defender of borderlineness principles can now appeal to the coherence, for all the paradoxes show, of a position which asserts borderlineness principles while keeping within precise boundaries its pattern of forgetfulness: if such a position is still coherent, no fatal threat would seem to be posed on borderlineness principles by the paradoxes of higher-order vagueness.^{35,36}

³⁵Of course, these are not counters available to a theory adopting the dominant approach, as they require acceptance of the radically deflationary view of borderline cases advocated in this chapter. Also, let me stress that, for any naive theory of vagueness, the last line of resistance against the paradoxes still remains their invalidity in tolerant logics.

³⁶True, on the proposed reconstruction of the notion, even though borderline cases of

The second difference is that, on the proposed reconstruction, the property of being a borderline case is not importantly determined by any intrinsic feature of the item in question and hence can vastly change from its being exemplified to its not being exemplified and *vice versa* according to the volatile vicissitudes of episodes of consideration. In some sense or other, it is *relative* to such episodes.³⁷ This is in sharp contrast with the interpretation given by the dominant approach, according to which the property of being a borderline case is importantly determined by intrinsic features of the item in question and hence is relatively invariant across episodes of consideration. This difference is just what one would have expected given what the source

a vague predicate will in many cases have precise boundaries, it may well be that the relevant propositions stating where these precise boundaries are on a certain episode of consideration e are forgotten on at least some episode of consideration. This will often be plausible especially on e itself: if one forgets about the question whether x is F , one often also forgets about the question whether x is borderline F . Thus, it may well be that the premises of the paradoxes of higher-order vagueness are still true on some episode of consideration. I accept this conclusion, but I am sceptical that anything problematic for my view follows from it. Firstly, on the present deflationary view of borderline cases, even if some paradox remained, it would *not* be a paradox of *vagueness* at all, but a paradox of *forgetfulness*, to be treated in one's favoured epistemological theory. Secondly, a straightforward solution to the alleged paradox does seem to be readily available, as the relevant closure principle for the property of being definite clearly stands no chance of being correct on the present deflationary view of definite cases. For example, on the present view, it can be the case that it is definite on an episode of consideration e that everything in the domain of discourse is F , but also that it is not definite on e that a certain member x of the domain is F , even though the latter follows from the former. This is so because it can be the case that, even though the proposition that everything is F is not forgotten on e , the proposition that x is F is forgotten on e (and so not definite on e).

³⁷There is by now a huge literature on the different possible semantic and metaphysical implementations of the intuitive idea that the correctness of the application of a notion is relative to some specified parameter. Since my proposal is reconstructive rather than descriptive, I don't think it's terribly important to decide here among these implementations, just as, *si parva licet*, this wasn't terribly important for special-relativity theorists with respect to the notion of simultaneity.

of borderline cases is on the view I am proposing: borderline cases arise not because some items have some distinctive property, but because a soritical series, even though finite in the classical sense, cannot be surveyed all at once by an inquirer,³⁸ and some item or other in it had better remain forgotten, lying in the penumbra of the light cast by inquiry. But it does not matter much *which* particular item so remains forgotten, and it is just natural that several items reach and leave the penumbra according to the different foci of attention of different episodes of consideration. On this view, the notion of a borderline case is peculiarly anthropomorphic and does not belong to the side of things as they are in themselves in abstraction from inquiry: the *esse* of borderline cases consists in their *(non) percipi*.

6.6 Conclusion

For the reasons I have already explained, the proposed conception of borderline cases is particularly hospitable to the theory of vagueness defended in this essay. From the standpoint of that theory, the overarching dialectic of this chapter looks as follows. The forced-fault paradox is particularly acute for a naive theory of vagueness, since, on any such theory, the drawing of sharp boundaries provoked by the forced march is objectively wrong (running afoul of tolerance principles) and not just unwarranted. Hence, a theory accepting tolerance principles is in even more urgent a need than other theories are of a satisfactory solution to the paradox. The forgetfulness solution is such a solution, being motivated independently of vague forced marches (as witnessed by precise forced marches) and indeed independently of forced marches in general (as witnessed by the Believer paradox). Crucially for the

³⁸The idea of totalities which, even though finite in the classical sense, are unsurveyable is central in the strict-finitist tradition (Hjelmslev [1922]; Esenin-Vol'pin [1970]; Vopěnka [1979] are some of its milestones). As I have warned in section 1.5, the exploration of the connections between this tradition and the position developed in this essay must wait for another occasion.

naive theory, the forgetfulness solution also lets a notion emerge which can serve as a viable reconstruction of the notion of a borderline case, since it can account for the phenomenon of borderlineness, which is arguably one of the few genuine phenomena to be found in this area. Tolerance and borderlineness are thus finally connected. In the face of the utter inability of the dominant approach to account for our sense that vague expressions do not draw sharp boundaries, the tie thus achieved seems to me another point in favour of the theory defended in this essay.

Chapter 7

Conclusion

This essay has offered a new theory of vagueness. The theory falls within the traditional approach to the problem of vagueness by taking as basic the phenomenon of sorites susceptibility. The main theoretical tasks of the theory are thus determined to be the explanation of this phenomenon, an account of its relation with sorites paradoxes and the explanation of the other phenomena of vagueness by means of the more basic property of sorites susceptibility. This essay can be seen as a sustained attempt at accomplishing these tasks for a specific theory which adopts the traditional approach.

The specific theory developed in this essay is a naive theory of vagueness, holding that the nature of the vagueness of a predicate consists in its failure to draw a sharp boundary between its positive and negative cases—that is, in its tolerance. I have argued in chapter 2 for the claim that tolerance is indeed a crucial feature that some of our concepts must possess if they are to serve in the achievement of certain important theoretical and practical purposes—if certain thoughts about, experiences of and interactions with the world are to be possible. Together with the very plausible assumption that the achievement of these purposes, if at all possible, is what vagueness is for, those arguments can be turned into arguments in favour of the naive theory itself, since rejecting the theory would then amount to rejecting that

the achievement of these purposes is at all possible—a sad conclusion indeed. From the standpoint of the naive theory, such arguments also contribute to an exhibition of the sources of vagueness. These sources are rooted in fundamental facts about our cognition and agency in the world. The naive theory thus gives vagueness a depth which is by far unmatched by all the other theories.

The truth of tolerance principles, together with the assumption that such truth is available to vague-concepts users, can also quite straightforwardly explain most phenomena of vagueness, such as sorites susceptibility, indefiniteness, ignorance and higher-order vagueness. At this point, the cumulative argument in favour of the naive theory may appear impressive: not only does the naive theory have an undeniable air of self-evidence about it and enjoy an enviable explanatory power, but it also connects in a desirable way vagueness with fundamental facts about our cognition, agency and experience in the world. This concurrence of positive features plausibly makes it the theory which must be defeated if it is not to be accepted.

Alas, there is indeed a defeater: the sorites paradox. As no other serious defeaters are in view, a satisfactory solution to the paradox becomes in effect the Grail of the naive theory of vagueness: if such a solution can be found, it will be hard to see, in view of its other positive features, how the theory could fail to be the right theory of vagueness. By way of a preliminary to the real treatment of the paradox, I have shown in chapter 3 that it will be in any event vain to retreat from the naive theory (and from its tolerance principles) to some weaker theory adopting the dominant approach (and to its borderlineness principles). For, once the full extent that higher-order vagueness must take under the dominant approach is recognized, it can at last be seen that the same kind of logic which, *via* a standard sorites paradox, dooms tolerance to inconsistency is also sufficient, *via* a higher-order sorites paradox, to doom borderlineness to inconsistency.

Coming to my own solution to the sorites paradox, I have identified in chapter 4 an unobvious usually hidden presupposition of the reasoning in-

volved in the paradox (the transitivity of the consequence relation) and developed a natural weakening of the logic which puts principled restrictions on it, thereby making the naive theory consistent. Some of the many philosophical issues arising from this non-transitivist solution to the sorites paradox have been addressed in chapter 5. There, I have argued that non-transitive conceptions of logical consequence quite generally find their roots in the idea that consequence should be faithful to certain local connections and that at least some of these conceptions can (even if they need not) be made sense of as representing a different understanding of the normative import of consequence rather than as contesting some traditional law linking consequence with truth.

Although the claim has not been adequately defended in this essay, I don't think that vagueness gives us any reason to believe in the existence of strong borderline cases. Yet, I do acknowledge the phenomenon of borderlineness, and so my theory had better say something about what it is to be a (non-strong) borderline case and why there are (non-strong) borderline cases. Indeed, there arguably is a general challenge for theories adopting the traditional approach to connect in a satisfactory way sorites susceptibility with borderlineness. I have tackled this issue in chapter 6, starting from a more general problem concerning the possibility and limits of the preservation of one's intellectual integrity in inquiry, and addressing from that—I hope—illuminating perspective the forced-march paradox. I have argued that the paradox is not really a paradox of vagueness at all and that no happy-face solution to it will succeed. I have proposed my own unhappy-face solution, which requires that one forget about some items in the forced-march series. Finally, exploiting the concept of a forgotten item crucially employed by this solution, I have suggested that, in a reconstructive spirit, we identify borderline cases on an episode of consideration with the relevant items which are forgotten on that episode of consideration and have argued that the reconstructed notion actually suffices to explain the phenomenon of borderlineness and other features of the pre-theoretical notion of a borderline case.

The riddle of vagueness has been a major cause of intellectual embarrassment for a few ancient philosophical schools and has occupied many of the sharpest minds in modern analytic philosophy. It lies at the crossroad of some of the deepest and hardest issues in the philosophy of logic, language and epistemology. I have really enjoyed thinking about it for the last few years and I must actually confess that I'm fairly confident that the theory presented in this essay is right in its main contentions. Yet, I think that, as is often the case for many of the perennial problems in philosophy, to manage to hit the truth is actually much less important for the intellectual value of our inquiry than to improve our understanding of what the problem really is, to assess merits and disadvantages of extant positions and to develop original ways of thinking about it. I don't think there can be any doubt that, on these latter dimensions, the contemporary debate on vagueness has been a success story of modern analytical philosophy. I hope that the present essay will fruitfully contribute to this ongoing dialogue.

Bibliography

- William Alston. *The Philosophy of Language*. Prentice Hall, Englewood Cliffs, 1964.
- David Armstrong. *A Materialist Theory of Mind*. Routledge, London, 1968.
- Arnon Avron. Simple consequence relations. *Information and Computation*, 92:105–40, 1991.
- Jonathan Barnes. Medicine, experience, and logic. In Jonathan Barnes, Jacques Brunschwig, Myles Burnyeat, and Malcolm Schoefield, editors, *Science and Speculation*, pages 24–68. Cambridge UP, Cambridge, 1982.
- David Barnett. Vagueness, knowledge, and rationality. URL http://spot.colorado.edu/~barnetdb/my_papers/Vagueness%20and%20Knowledge.pdf. ms, 2007a.
- David Barnett. Is vagueness sui generis? URL http://spot.colorado.edu/~barnetdb/my_papers/Barnett_Vagueness.pdf. ms, 2007b.
- Jon Barwise and John Perry. *Situations and Attitudes*. MIT Press, Cambridge MA, 1983.
- Jean-Yves Béziau. Transitivity and paradoxes. In Jurgis Škilters, Matti Eklund, Ólafur Páll Jónsson, and Olav Wiegand, editors, *Paradox: Logical, Cognitive and Communicative Aspects*, volume I of *The Baltic International Yearbook of Cognition, Logic and Communication*, pages 87–92. University of Latvia Press, Riga, 2006.

- Susanne Bobzien. Chrysippus and the epistemic theory of vagueness. *Proceedings of the Aristotelian Society*, 102:217–38, 2002.
- Bernhard Bolzano. *Wissenschaftslehre*. Seidel, Sulzbach, 1837.
- Robert Brandom. The significance of complex numbers for Frege’s philosophy of mathematics. *Proceedings of the Aristotelian Society*, 96:293–315, 1996.
- David Braun and Theodore Sider. Vague, so untrue. *Nous*, 41:133–56, 2007.
- Wylie Breckenridge. *The Meaning of ‘Look’*. PhD thesis, Faculty of Philosophy, Oxford, 2007. URL <http://users.ox.ac.uk/~newc1666/DPhilThesis.pdf>.
- John Broome. Normative requirements. *Ratio*, 12:398–419, 1999.
- Tyler Burge. Epistemic paradox. *The Journal of Philosophy*, 81:5–29, 1984.
- John Burgess and Lloyd Humberstone. Natural deduction rules for a logic of vagueness. *Erkenntnis*, 27:197–229, 1987.
- Linda Burns. Vagueness and coherence. *Synthese*, 68:487–513, 1986.
- Linda Burns. *Vagueness*. Kluwer, Dordrecht, 1991.
- Linda Burns. Something to do with vagueness. *Southern Journal of Philosophy Supplement*, 33:23–47, 1995.
- Myles Burnyeat. Gods and heaps. In Malcolm Schofield and Martha Nussbaum, editors, *Language and Logos*, pages 315–38. Cambridge UP, Cambridge, 1982.
- Roderick Chisholm. *Perceiving: A Philosophical Study*. Cornell UP, Ithaca, 1957.
- David Christensen. *Putting Logic in its Place*. Oxford UP, Oxford, 2004.

- Pablo Cobreros. Supervaluations and logical consequence: Retrieving the local perspective. URL <http://www.unav.es/filosofia/pcobrereros/Spv%20and%20LC.pdf>. ms, 2007.
- Roy Cook. Clarifying the logic of clear cases. In Sebastiano Moruzzi and Andrea Sereni, editors, *Issues on Vagueness*, pages 75–104. Il Poligrafo, Padova, 2005.
- Jack Copeland. Vague identity and fuzzy logic. *The Journal of Philosophy*, 94:514–34, 1997.
- Cian Dorr. Vagueness without ignorance. *Philosophical Perspectives*, 17: 83–113, 2003.
- Michael Dummett. Wang’s paradox. *Synthese*, 30:301–24, 1975a.
- Michael Dummett. The justification of deduction. In *Truth and Other Enigmas*, pages 290–318. Duckworth, London, 1975b.
- Michael Dummett. Common sense and physics. In Graham MacDonald, editor, *Perception and Identity*, pages 1–40. Macmillan, London, 1979.
- Michael Dummett. *The Interpretation of Frege’s Philosophy*. Duckworth, London, 1981.
- Catarina Dutilh-Novaes. *Formalizing Medieval Logical Theories*. Springer, New York, 2007.
- Dorothy Edgington. Validity, uncertainty and vagueness. *Analysis*, 52:193–204, 1992.
- Dorothy Edgington. Wright and Sainsbury on higher-order vagueness. *Analysis*, 53:193–200, 1993.
- Matti Eklund. Inconsistent languages. *Philosophy and Phenomenological Research*, 64:251–75, 2001.

- Matti Eklund. What vagueness consists in. *Philosophical Studies*, 125:27–60, 2005.
- Richard Epstein. Relatedness and implication. *Philosophical Studies*, 36: 137–73, 1979.
- Aleksandr Esenin-Vol’pin. The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics. In Akiko Kino, John Myhill, and Richard Vesley, editors, *Intuitionism and Proof Theory*, pages 3–45. North-Holland, Amsterdam, 1970.
- Gareth Evans. Semantic structure and logical form. In Gareth Evans and John McDowell, editors, *Truth and Meaning*, pages 199–222. Oxford UP, Oxford, 1978.
- Delia Graff Fara. Shifting sands: An interest-relative theory of vagueness. *Philosophical Topics*, 28:45–81, 2000.
- Delia Graff Fara. Phenomenal continua and the sorites. *Mind*, 110:905–35, 2001.
- Delia Graff Fara. An anti-epistemicist consequence of margin for error semantics for knowledge. *Philosophy and Phenomenological Research*, 64: 127–42, 2002.
- Delia Graff Fara. Gap principles, penumbral consequence, and infinitely higher-order vagueness. In JC Beall, editor, *Liars and Heaps*, pages 195–221. Oxford UP, Oxford, 2003.
- Hartry Field. Disquotational truth and factually defective discourse. *The Philosophical Review*, 103:405–52, 1994.
- Hartry Field. Indeterminacy, degree of belief, and excluded middle. *Nous*, 34:1–30, 2000.
- Hartry Field. Saving the truth schema from paradox. *Journal of Philosophical Logic*, 31:1–27, 2002.

- Hartry Field. The semantic paradoxes and the paradoxes of vagueness. In JC Beall, editor, *Liars and Heaps*, pages 262–311. Oxford UP, Oxford, 2003a.
- Hartry Field. No fact of the matter. *Australasian Journal of Philosophy*, 81: 457–80, 2003b.
- Hartry Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32:139–77, 2003c.
- Hartry Field. Truth and the unprovability of consistency. *Mind*, 115:567–605, 2006.
- Hartry Field. Solving the paradoxes, escaping revenge. URL <http://philosophy.fas.nyu.edu/docs/I0/1158/revengetex3.pdf>. ms, 2007a.
- Hartry Field. Vagueness, partial belief, and logic. URL <http://philosophy.fas.nyu.edu/docs/I0/1158/schiffer2004b.pdf>. ms, 2007b.
- Kit Fine. Vagueness, truth and logic. *Synthese*, 30:265–300, 1975.
- Roderick Firth. Are epistemic concepts reducible to moral concepts? In Alvin Goldman and Jaegwon Kim, editors, *Values and Morals*, pages 215–29. Reidel, Dordrecht, 1978.
- Gottlob Frege. *Grundgesetze der Arithmetik*, volume I. Pohle, Jena, 1893.
- Gottlob Frege. *Grundgesetze der Arithmetik*, volume II. Pohle, Jena, 1903.
- Haim Gaifman. Vagueness, tolerance and contextual logic. URL <http://www.columbia.edu/~hg17/VTC-latest.pdf>. ms, 2007.
- Manuel García-Carpintero. The grounds for the model-theoretic account of the logical properties. *Notre Dame Journal of Formal Logic*, 34:107–31, 1993.
- Peter Geach. Entailment. *Proceedings of the Aristotelian Society Supplementary Volume*, 32:157–72, 1958.

- Rolf George. Bolzano's consequence, relevance, and enthymemes. *Journal of Philosophical Logic*, 12:299–318, 1983.
- Rolf George. Bolzano's concept of consequence. *The Journal of Philosophy*, 83:558–64, 1986.
- Mario Gómez-Torrente. Two problems for an epistemicist view of vagueness. *Philosophical Issues*, 8:237–45, 1997.
- Mario Gómez-Torrente. The problem of logical constants. *The Bulletin of Symbolic Logic*, 8:1–37, 2002a.
- Mario Gómez-Torrente. Vagueness and margin for error principles. *Philosophy and Phenomenological Research*, 64:107–25, 2002b.
- Nelson Goodman. *The Structure of Appearance*. Harvard UP, Cambridge MA, 1951.
- Patrick Greenough. Vagueness: A minimal theory. *Mind*, 112:235–81, 2003.
- Patrick Greenough. Contextualism about vagueness and higher-order vagueness. *Proceedings of the Aristotelian Society Supplementary Volume*, 79:167–90, 2005.
- Clyde Hardin. Phenomenal colors and sorites. *Nous*, 22:213–34, 1988.
- Gilbert Harman. *Change in View*. MIT Press, Cambridge MA, 1986.
- John Hawthorne. *Knowledge and Lotteries*. Oxford UP, Oxford, 2004.
- John Hawthorne. Vagueness and the mind of God. *Philosophical Studies*, 122:1–25, 2005.
- Richard Heck. A note on the logic of (higher-order) vagueness. *Analysis*, 53:201–8, 1993.
- Richard Heck. Semantic accounts of vagueness. In JC Beall, editor, *Liars and Heaps*, pages 106–27. Oxford UP, Oxford, 2003.

Johannes Hjeltslev. Die natürliche Geometrie. *Abhandlungen aus dem Mathematischen Seminar der Hamburgischen Universität*, 2:1–36, 1922.

Terence Horgan. Robust vagueness and the forced-march sorites paradox. *Philosophical Perspectives*, 8:159–88, 1994.

Dominic Hyde. From heaps and gaps to heaps of gluts. *Mind*, 106:641–60, 1997.

Dominic Hyde and Richard Sylvan. Ubiquitous vagueness without embarrassment. *Acta Analytica*, 10:7–29, 1993.

Frank Jackson and Ray Pinkerton. On an argument against sensory items. *Mind*, 82:269–72, 1973.

Hans Kamp. Two theories about adjectives. In Edward Keenan, editor, *Formal Semantics of Natural Language*, pages 123–55. Cambridge UP, Cambridge, 1975.

Hans Kamp. The paradox of the heap. In Uwe Mönnich, editor, *Aspects of Philosophical Logic*, pages 225–77. Reidel, Dordrecht, 1981.

David Kaplan and Richard Montague. A paradox regained. *Notre Dame Journal of Formal Logic*, 1:79–90, 1960.

Rosanna Keefe. *Theories of Vagueness*. Cambridge UP, Cambridge, 2000.

Rosanna Keefe. Context, vagueness, and the sorites: Comments on Shapiro. In JC Beall, editor, *Liars and Heaps*, pages 73–83. Oxford UP, Oxford, 2003.

Rosanna Keefe. Vagueness without context change. *Mind*, 116:275–92, 2007.

Rosanna Keefe and Peter Smith. Introduction. In Rosanna Keefe and Peter Smith, editors, *Vagueness: A Reader*, pages 1–57. MIT Press, Cambridge MA, 1997.

- Keith Korcz. Recent work on the basing relation. *American Philosophical Quarterly*, 34:171–89, 1997.
- Kazimierz Kuratowski. Sur l'opération \overline{A} de l'analysis situs. *Fundamenta Mathematicae*, 3:182–99, 1922.
- George Lakoff. Hedges: A study in meaning criteria and the logic of fuzzy concepts. *Journal of Philosophical Logic*, 2:458–508, 1973.
- David Lewis. General semantics. *Synthese*, 22:189–229, 1970.
- Casimir Lewy. Entailment. *Proceedings of the Aristotelian Society Supplementary Volume*, 32:123–42, 1958.
- Casimir Lewy. *Meaning and Modality*. Cambridge UP, Cambridge, 1976.
- Bernard Linsky. Phenomenal qualities and the identity of indistinguishables. *Synthese*, 59:363–80, 1984.
- Jonathan Lowe. Vague identity and quantum indeterminacy. *Analysis*, 54:110–4, 1994.
- David Makinson. The paradox of the preface. *Analysis*, 25:205–7, 1965.
- David Makinson. *Bridges from Classical to Nonmonotonic Logic*. King's College London Publications, London, 2005.
- Erroll Martin and Robert Meyer. S (for syllogism). ms, 1982.
- Paul Masson-Oursel. Esquisse d'une théorie comparée du sorite. *Revue de métaphysique et de morale*, 20:810–24, 1912.
- Vann McGee. *Truth, Vagueness, and Paradox*. Hackett, Indianapolis, 1991.
- Vann McGee. Two problems with Tarski's theory of consequence. *Proceedings of the Aristotelian Society*, 92:273–92, 1992.
- Vann McGee and Brian McLaughlin. Distinctions without a difference. *Southern Journal of Philosophy Supplement*, 33:204–51, 1995.

- Mario Mignucci. The stoic analysis of the sorites. *Proceedings of the Aristotelian Society*, 93:231–45, 1993.
- Jon Moline. Aristotle, Eubulides, and the sorites. *Mind*, 78:393–407, 1969.
- Sebastiano Moruzzi and Elia Zardini. Conseguenza logica. In Annalisa Coliva, editor, *Filosofia analitica*, pages 157–94. Carocci, Roma, 2007.
- Peter Mott. Margins for error and the sorites paradox. *The Philosophical Quarterly*, 48:494–504, 1998.
- Francesco Paoli. *Substructural Logics: A Primer*. Kluwer, Dordrecht, 2002.
- Rohit Parikh. Existence and feasibility in arithmetic. *The Journal of Symbolic Logic*, 36:494–508, 1971.
- Rohit Parikh. The problem of vague predicates. In Robert Cohen and Marx Wartofsky, editors, *Language, Logic, and Method*, volume 31 of *Boston Studies in the Philosophy of Science*, pages 241–61. Reidel, Dordrecht, 1983.
- Terence Parsons. Assertion, denial, and the liar paradox. *Journal of Philosophical Logic*, 13:137–52, 1984.
- Cristopher Peacocke. Are vague predicates incoherent? *Synthese*, 46:121–41, 1981.
- Karl Popper. *The Logic of Scientific Discovery*. Basic Books, New York, 1959.
- Graham Priest. Sorites and identity. *Logique et Analyse*, 135:293–6, 1991.
- Graham Priest. Fuzzy identity and local validity. *The Monist*, 81:331–42, 1998.
- Graham Priest. *An Introduction to Non-Classical Logic*. Cambridge UP, Cambridge, 2001.

- Graham Priest. A site for sorites. In JC Beall, editor, *Liars and Heaps*, pages 9–23. Oxford UP, Oxford, 2003.
- Graham Priest. *In Contradiction*. Oxford UP, Oxford, 2nd edition, 2006a.
- Graham Priest. *Doubt Truth to Be a Liar*. Oxford UP, Oxford, 2006b.
- Arthur Prior. *Objects of Thought*. Clarendon Press, Oxford, 1971.
- Jim Pryor. What’s wrong with Moore’s argument? *Philosophical Issues*, 14: 349–78, 2004.
- Willard Quine. *Word and Object*. MIT Press, Cambridge MA, 1960.
- Willard Quine. *Philosophy of Logic*. Harvard UP, Cambridge MA, 2nd edition, 1986.
- Diana Raffman. Vagueness without paradox. *The Philosophical Review*, 103: 41–74, 1994.
- Diana Raffman. Vagueness and context-relativity. *Philosophical Studies*, 81: 175–92, 1996.
- Diana Raffman. Is perceptual indiscriminability nontransitive? *Philosophical Topics*, 28:153–75, 2000.
- Stig Rasmussen. Well-ordering and the sorites paradox. *Danish Yearbook of Philosophy*, 23:59–119, 1986.
- Stephen Read. Validity and the intensional sense of ‘and’. *Australasian Journal of Philosophy*, 59:301–7, 1981.
- Stephen Read. *Relevant Logic*. Blackwell, Oxford, 1988.
- Stephen Read. Formal and material consequence. *Journal of Philosophical Logic*, 23:247–65, 1994.
- Stephen Read. Logical consequence as truth-preservation. *Logique et Analyse*, 183:479–93, 2003.

- Alfréd Rényi. *Foundations of Probability*. Holden-Day, San Francisco, 1970.
- Greg Restall. *An Introduction to Substructural Logics*. Routledge, London, 2000.
- Teresa Robertson. On Soames's solution to the sorites paradox. *Analysis*, 60:328–34, 2000.
- Bertil Rolf. *Topics on Vagueness*. Studentlitteratur, Lund, 1981.
- Bertil Rolf. Sorites. *Synthese*, 58:219–50, 1984.
- Richard Routley, Val Plumwood, Robert Meyer, and Ross Brady. *Relevant Logics and Their Rivals*, volume I. Ridgeview, Atascadero, 1982.
- Ian Rumfitt. “Yes” and “No”. *Mind*, 109:781–823, 2000.
- Bertrand Russell. *The Analysis of Matter*. Kegan Paul, London, 1927.
- Bertrand Russell. *An Inquiry into Meaning and Truth*. Allen and Unwin, London, 1950.
- Mark Sainsbury. Tolerating vagueness. *Proceedings of the Aristotelian Society*, 89:33–48, 1989.
- Mark Sainsbury. Concepts without boundaries. In Rosanna Keefe and Peter Smith, editors, *Vagueness: A Reader*, pages 251–64. MIT Press, Cambridge MA, 1990.
- Mark Sainsbury. Is there higher-order vagueness? *The Philosophical Quarterly*, 41:67–82, 1991.
- Mark Sainsbury. Sorites paradoxes and the transition question. *Philosophical Papers*, 21:177–90, 1992.
- Mark Sainsbury. *Paradoxes*. Cambridge UP, Cambridge, 2nd edition, 1995.
- Mark Sainsbury. *Logical Forms*. Blackwell, Oxford, 2nd edition, 2001.

- David Sanford. Borderline logic. *American Philosophical Quarterly*, 12:29–39, 1975.
- Stephen Schiffer. Two issues of vagueness. *The Monist*, 88:193–214, 1998.
- Stephen Schiffer. Vagueness and partial belief. *Philosophical Issues*, 10:220–57, 2000.
- Stephen Schiffer. *The Things We Mean*. Oxford UP, Oxford, 2003.
- Dana Scott. Completeness and axiomatizability in many-valued logic. In Leon Henkin, editor, *Proceedings of the Tarski Symposium*, pages 411–35. American Mathematical Society, Providence RI, 1974.
- Stewart Shapiro. Vagueness and conversation. In JC Beall, editor, *Liars and Heaps*, pages 39–72. Oxford UP, Oxford, 2003.
- Stewart Shapiro. *Vagueness in Context*. Oxford UP, Oxford, 2006.
- Giovanna Sillitti. Aristotele e l’aporia del sorite. *Elenchos*, 5:211–20, 1984.
- Timothy Smiley. Entailment and deducibility. *Proceedings of the Aristotelian Society*, 59:233–54, 1959.
- Timothy Smiley. Rejection. *Analysis*, 56:1–9, 1996.
- Nicholas Smith. Vagueness and blurry sets. *Journal of Philosophical Logic*, 33:165–235, 2004.
- Scott Soames. *Understanding Truth*. Oxford UP, Oxford, 1999.
- Roy Sorensen. *Blindspots*. Clarendon Press, Oxford, 1988.
- Roy Sorensen. The ambiguity of vagueness and precision. *Pacific Philosophical Quarterly*, 70:174–83, 1989.
- Roy Sorensen. Ambiguity, discretion, and the sorites. *The Monist*, 81:217–35, 1998.

- Roy Sorensen. *Vagueness and Contradiction*. Oxford UP, Oxford, 2001.
- Roy Sorensen. Knowledge beyond the margin for error. *Mind*, 116:717–22, 2007.
- Jason Stanley. Context, interest-relativity, and the sorites. *Analysis*, 63: 269–80, 2003.
- Richard Sylvan. *Sociative Logics and their Applications*. Ashgate, Aldershot, 2000.
- Jamie Tappenden. The liar and sorites paradoxes: Toward a unified treatment. *The Journal of Philosophy*, 90:551–77, 1993.
- Jamie Tappenden. Negation, denial and language change in philosophical logic. In Dov Gabbay and Heinrich Wansing, editors, *What Is Negation?*, pages 261–98. Kluwer, Dordrecht, 1999.
- Alfred Tarski. Fundamental concepts of the methodology of deductive sciences. In *Logic, Semantics, Metamathematics*, pages 60–109. Hackett, Indianapolis, 2nd edition, 1930.
- Alfred Tarski. On the concept of logical consequence. In *Logic, Semantics, Metamathematics*, pages 409–20. Hackett, Indianapolis, 2nd edition, 1936.
- Neil Tennant. *Anti-Realism and Logic*. Clarendon Press, Oxford, 1987.
- Charles Travis. Vagueness, observation, and sorites. *Mind*, 94:345–66, 1985.
- Michael Tye. Vague objects. *Mind*, 99:535–57, 1990.
- Michael Tye. Sorites paradoxes and the semantics of vagueness. *Philosophical Perspectives*, 8:189–206, 1994.
- Peter Unger. I do not exist. In Graham MacDonald, editor, *Perception and Identity*, pages 235–51. Macmillan, London, 1979a.
- Peter Unger. There are no ordinary things. *Synthese*, 41:117–54, 1979b.

- Peter Unger. Why there are no people. *Midwest Studies in Philosophy*, 4: 177–222, 1979c.
- Peter Unger. Scepticism and nihilism. *Nous*, 14:517–45, 1980.
- Jean van Heijenoort. Frege and vagueness. In Leila Haaparanta and Jaakko Hintikka, editors, *Frege Synthesized*, pages 31–45. Reidel, Dordrecht, 1986.
- Achille Varzi. Supervaluationism and paraconsistency. In Diderik Batens, Chris Mortensen, Graham Priest, and Jean-Paul van Bendegem, editors, *Frontiers in Paraconsistent Logic*, pages 279–97. Research Studies Press, Baldock, 2000.
- Achille Varzi. On logical relativity. *Philosophical Issues*, 10:197–219, 2002.
- Georg von Wright. The concept of entailment. In *Logical Studies*, pages 166–91. Routledge, London, 1957.
- Petr Vopěnka. *Mathematics in the Alternative Set Theory*. Teubner, Leipzig, 1979.
- Douglas Walton. Philosophical basis of relatedness logic. *Philosophical Studies*, 36:115–36, 1979.
- Brian Weatherson. True, truer, truest. *Philosophical Studies*, 123:47–70, 2005a.
- Brian Weatherson. Vague terms with sharp boundaries: Reply to Greenough. URL <http://brian.weatherson.org/cpg.pdf>. ms, 2005b.
- Brian Weatherson. Vagueness as indeterminacy. URL <http://brian.weatherson.org/vai-oct.pdf>. ms, 2006.
- Ruth Weintraub. On sharp boundaries for vague terms. *Synthese*, 138:233–45, 2004.
- Alan Weir. Naïve set theory is innocent! *Mind*, 107:763–98, 1998.

- Stephen Weiss. The sorites fallacy: What difference does a peanut make? *Synthese*, 33:253–72, 1976.
- Samuel Wheeler. Reference and vagueness. *Synthese*, 30:367–79, 1975.
- Samuel Wheeler. On that which is not. *Synthese*, 41:155–94, 1979.
- Timothy Williamson. *Identity and Discrimination*. Blackwell, Oxford, 1990.
- Timothy Williamson. Inexact knowledge. *Mind*, 101:217–42, 1992.
- Timothy Williamson. *Vagueness*. Routledge, London, 1994.
- Timothy Williamson. Cognitive homelessness. *The Journal of Philosophy*, 93:554–73, 1996.
- Timothy Williamson. Replies to commentators. *Philosophical Issues*, 8:255–65, 1997.
- Timothy Williamson. On the structure of higher-order vagueness. *Mind*, 108: 127–43, 1999.
- Timothy Williamson. Margin for error: A reply. *The Philosophical Quarterly*, 50:76–81, 2000a.
- Timothy Williamson. *Knowledge and its Limits*. Oxford UP, Oxford, 2000b.
- Timothy Williamson. Epistemicist models: Comments on Gómez-Torrente and Graff. *Philosophy and Phenomenological Research*, 64:143–50, 2002.
- Timothy Williamson. Knowledge within the margin for error. *Mind*, 116: 723–6, 2007.
- Ryszard Wójcicki. *Theory of Logical Calculi. Basic Theory of Consequence Operations*. Kluwer, Dordrecht, 1988.
- Crispin Wright. On the coherence of vague predicates. *Synthese*, 30:325–65, 1975.

- Crispin Wright. Language mastery and the sorites paradox. In Gareth Evans and John McDowell, editors, *Truth and Meaning*, pages 223–47. Oxford UP, Oxford, 1976.
- Crispin Wright. Further reflections on the sorites paradox. *Philosophical Topics*, 15:227–90, 1987.
- Crispin Wright. The sorites paradox and its significance for the interpretation of semantic theory. In Neil Cooper and Pascal Engel, editors, *New Inquiries Into Meaning and Truth*, pages 135–62. Harvester Wheatsheaf, Hemel Hempsted, 1991.
- Crispin Wright. Is higher-order vagueness coherent? *Analysis*, 52:129–39, 1992.
- Crispin Wright. On being in a quandary: Relativism, vagueness, logical revisionism. *Mind*, 60:45–98, 2001.
- Crispin Wright. Vagueness: A fifth column approach. In JC Beall, editor, *Liars and Heaps*, pages 84–105. Oxford UP, Oxford, 2003a.
- Crispin Wright. Rosenkranz on quandary, vagueness and intuitionism. *Mind*, 112:465–74, 2003b.
- Crispin Wright. On the characterisation of borderline cases. URL <http://philosophy.fas.nyu.edu/docs/I0/1187/borderline.pdf>. ms, 2007a.
- Crispin Wright. The illusion of higher-order vagueness. ms, 2007b.
- Crispin Wright. Wang’s paradox. URL <http://philosophy.fas.nyu.edu/docs/I0/1187/dummett.pdf>. ms, 2007c.
- Lofti Zadeh. Fuzzy logic and approximate reasoning. *Synthese*, 30:407–28, 1975.
- Elia Zardini. Squeezing and stretching: How vagueness can outrun borderlineness. *Proceedings of the Aristotelian Society*, 106:419–26, 2006a.

Elia Zardini. Higher-order vagueness and paradox: The glory and misery of S4 definiteness. In Jurgis Škilters, Matti Eklund, Ólafur Páll Jónsson, and Olav Wiegand, editors, *Paradox: Logical, Cognitive and Communicative Aspects*, volume I of *The Baltic International Yearbook of Cognition, Logic and Communication*, pages 203–20. University of Latvia Press, Riga, 2006b.

Elia Zardini. Red and orange. ms, 2007a.

Elia Zardini. Sequent calculi for tolerant logics. ms, 2007b.

Elia Zardini. Logical consequence and the preservation of truth. ms, 2007c.

Elia Zardini. Inexact knowledge, transparency, and closure. ms, 2007d.

Paul Ziff. The number of English sentences. *Foundations of Language*, 11: 519–32, 1974.

Paul Ziff. *Epistemic Analysis*. Reidel, Dordrecht, 1984.