DOCTORAL THESIS

NATURAL LANGUAGE CONDITIONALS

(LOS CONDICIONALES DEL LENGUAJE NATURAL)

Fdo.: Martin Andor

DOUTORAMENTO EN LÓXICA E FILOSOFÍA DA CIENCIA

FACULTADE DE FILOSOFÍA

SANTIAGO DE COMPOSTELA

2015

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Departamento de Lóxica e Filosofia Moral
Praza de Mazarelos s/n
15782 Santiago de Compostela

Concepción Martínez Vidal, Profesora Titular del área de Lógica y Filosofía de la Ciencia en el Departamento de Lógica y Filosofía Moral de la USC

INFORMA QUE

La presente tesis, "Natural Language Conditionals" de la cual es autor Martin Andor, cumple todos los requisitos exigidos por la legislación vigente

Por eso, como tutora suya, firmo el correspondiente VISTO BUENO en Santiago de Compostela, a 22 de abril de 2015

Concepción Martínez Vidal
Profesora Titular Área de Lógica y Fil. Ciencia
Departamento de Lógica y Filosofía Moral
Universidad de Santiago de Compostela

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Universidad de Granada

Departamento de Filosofía I

María José Frápolli Sanz, Catedrática de Lógica y Filosofía de la Ciencia en el Departamento de Filosofía I de la Universidad de Granada,

PONGO EN CONOCIMIENTO DE QUIEN CORRESPONDA

Que la presente tesis, "Natural Language Conditionals", de la cual es autor Martin Andor, cumple todos los requisitos exigidos por la legislación vigente.

Por eso, como co-directora del trabajo, firmo el correspondiente VISTO BUENO en Londres, a 23 de abril de 2015

Fdo. Mª José Frápolli Sanz

Facultad de Filosofía y Letras - Edificio de Psicología - Campus de Cartuja - 18071 Granada, España
Teléfono: 958 243785 - Fax: 958 248981 – filosofi@ugr.es

Summaries

Resumo

Nesta tese doutoral, que se insire na tradición analítica da filosofía de linguaxe contemporánea, enfróntome a algúns dos temas máis discutidos en relación coa semántica dos condicionais da linguaxe natural. Tento ofrecer unha teoría plausíbel do seu significado, que sexa quen non só de dar unha solución aos crebacabezas clásicos (os paradoxos do condicional material e estrito, os problemas de Goodman e Tichý, as diferenzas entre os indicativos e os subxuntivos, etc.), senón tamén de explicar o seu xurdimento.

Argumento que as aproximacións veritativo-condicionais ao significado frecuentemente se cimentan en supostos implícitos inxustificados e adoitan conducir ao reducionismo semántico. Porén, non hai ningunha razón para pensar que algunhas das nosas expresións mais sofisticadas deban ser reducíbeis a algún vocabulario presuntamente máis elemental. En efecto, moitas tentativas de proporcionar unha análise veritativo-condicional de oracións condicionais resultaron flagrantemente circulares.

Un xeito máis natural de explicar o significado dos condicionais consiste en esclarecer a función que levan a cabo en intercambios lingüísticos ordinarios. Dende esta perspectiva, as oracións condicionais indicativas preséntanse como instrumentos para confinar a validez do acto de fala realizado por medio da oración principal. Iso permítenos explicar tanto os paradoxos do condicional material como a súa utilidade na matemática, igual que unha serie de fenómenos tales como os condicionais 'biscuit' e o papel dos condicionais na restrición cuantificacional.

No que respecta aos condicionais subxuntivos, eles son usados para realizaren afirmacións simples acerca do potencial inferencial do seu antecedente con respecto a un escenario inferencial ben delimitado. Argumento que a nosa concepción do mundo proporciona xusto un escenario así: o que sabemos da realidade organizámolo en padróns que frecuentemente posúen unha natureza causal. Son estes padróns os que sustentan o noso uso cotián de contrafácticos.

Palabras chave: condicionais, linguaxe natural, lóxica, probabilidade, causalidade.

Resumen

En esta disertación doctoral, que se inserta en la tradición analítica de la filosofía de lenguaje contemporánea, me enfrento a algunos de los temas más discutidos en relación con la semántica de los condicionales del lenguaje natural. Procuro ofrecer una teoría plausible de su significado, capaz no solo de dar una solución a los rompecabezas clásicos (paradojas del condicional material y estricto, los problemas de Goodman y Tichý, las diferencias entre los indicativos y subjuntivos, etc.), sino también de explicar su surgimiento.

Argumento que las aproximaciones veritativo-condicionales al significado frecuentemente se cimientan en supuestos implícitos injustificados y suelen llevar al reduccionismo semántico. Ahora bien, no hay ninguna razón para pensar que algunas de nuestras expresiones más sofisticadas deban ser reducibles a algún vocabulario presuntamente más elemental. En efecto, muchas tentativas de proporcionar un análisis veritativo-condicional de oraciones condicionales han resultado flagrantemente circulares.

Una manera más natural de explicar el significado de los condicionales consiste en esclarecer la función que llevan a cabo en intercambios lingüísticos ordinarios. Desde esta perspectiva, las oraciones condicionales indicativas se presentan como instrumentos para limitar la validez del acto de habla realizado por medio de la oración principal. Eso nos permite explicar tanto las paradojas del condicional material como su utilidad en la matemática, al igual que una serie de fenómenos tales como los condicionales 'biscuit' y el papel de los condicionales en la restricción cuantificacional.

Los condicionales subjuntivos, por su parte, realizan afirmaciones simples acerca del potencial inferencial de su antecedente con respecto a un escenario inferencial bien definido. Argumento que nuestra concepción del mundo proporciona justo un escenario así: lo que sabemos de la realidad lo organizamos en patrones, frecuentemente de naturaleza causal. Son estos patrones los que sustentan nuestro uso cotidiano de contrafácticos.

Palabras clave: condicionales, lenguaje natural, lógica, probabilidad, causalidad.

Abstract

In this PhD dissertation, set within the analytic tradition of the contemporary philosophy of language, I take on some of the most debated issues concerning the semantics of natural language conditionals. I strive to give a plausible account of their meaning, which would not only ward off the classical puzzles (paradoxes of the material and strict conditional, Goodman's and Tichý's problems, differences between indicatives and subjunctives, etc.), but also explain why they arise in the first place.

I argue that truth-conditional approaches to meaning are often based on unwarranted implicit assumptions and tend to favour semantic reductionism. Yet there is no reason for us to expect that some of our most sophisticated expressions should be logically reducible to any purportedly more basic vocabulary. And indeed, many attempts at a truth-conditional analysis of conditional clauses have turned out to be blatantly vacuous.

A more natural way to explain the meaning of conditionals is to point out the function they carry out in standard language exchanges. Approached from this perspective, indicative conditional clauses are arguably best viewed as devices limiting the validity of the speech act performed by the main clause. This offers us a handle on both the paradoxes of the material conditional and its usefulness in mathematics, as well as a range of other phenomena, such as the so-called biscuit conditionals and quantification restriction.

Subjunctive conditionals, on the other hand, make simple assertions about the inferential potential of the hypotheses introduced by their antecedents, given a well-defined inferential scenario. I argue that our picture of the world provides such a scenario—there are few bare facts, as almost all that we know about the reality is organised in (roughly causal) patterns. It is these patterns that support our everyday assertions of counterfactuals.

Key words: conditionals, natural language, logic, probability, causality.

Acknowledgements

First and foremost, I am unspeakably grateful to my supervisor, María José Frápolli Sanz, for her unrelenting support and for the freedom she has given me in my research (and which I have for the most part squandered). My co-supervisor Concepción Martínez Vidal has always been there for me and her help with administrative tasks has been invaluable.

I am thankful to my parents for always letting me make the mistakes I needed to in order to grow up, even though it was against their better judgement.

Thanks to Maristela, who always manages to bring a little more sunlight into my day.

Many thanks also to all my marvellous friends and colleagues, whose merits are too vast to be all mentioned here.

And last but not least, my gratitude goes to the hard-pressed Spanish taxpayers, who have financed my research despite undoubtedly having better ideas on how to spend their money in this period of crisis. I hope to repay them one day.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Contents

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 1

Introduction

'Yet another dissertation on conditionals.' If this was the first thing that crossed your mind upon seeing the title of this essay, you're excused—the topic has really taken off within analytical philosophy since Frege came up with his ingenious truth-functional analogue of the common 'if'. Trying to keep up with the rapidly growing literature on the subject resembles chasing a snowball down a snowy slope. It is therefore no surprise that most reasonable people have for all effects and purposes given up. Theorising about conditionals has taken a shape that we know from cancer research—almost everyone seems to have carved up his or her own niche topic, rarely feeling competent enough to venture to comment on the work going on in other sub-areas of the field.

You are excused even if that first idea of yours came accompanied with a sigh. I am afraid that unlike with cancer, the proliferation of philosophical research into conditionals has produced disproportionately meagre results. The chances that a new publication on the topic will contain a stunning original insight, instead of some clever, but otherwise insignificant tweak of an existing semantics, are admittedly slim. Intangible concepts, such as 'similarity ordering', 'contextual salience' or 'conversational background' abound, preventing any straightforward application of the intricate theories to real language phenomena. It is quite revealing that when Tetlock and Belkin tried to glean some wisdom for their work on counterfactual reasoning in history from the most popular philosophical theory on the subject, they were left practically empty-handed.[1]

One reason for the current situation may be a publishing culture that rewards neatly constructed articles focusing on very limited issues. This approach doesn't seem to do philosophy much good—she is much better

[1](Tetlock and Belkin, 1996), p.18, fn. 2

at broad strokes while leaving the details for special sciences to articulate. People have come to expect her to lead the way, not to improve on the navigational instruments once the course is set. The more technical a piece of philosophical writing, the more likely it is to pass unnoticed by anyone who might actually use it for some independent purpose. And when philosophical theorising becomes an end in itself, it tends to get out of hand.

But perhaps an even more important problem is that we seem to have lost a clear idea of the actual goals of our theoretical endeavour. Are we trying to elucidate the *meaning* of natural language conditionals? Are we aiming to specify the circumstances that must obtain for a conditional to hold? Should we guide ourselves by the way speakers use conditionals in everyday conversational exchanges? And is there a difference between these projects at all? Of course, each one of the mentioned theoretical challenges deserves the attention of philosophers. However, before taking the plunge, we should reserve some time to reflect upon the ultimate aims of our analysis and ponder whether the method we have chosen is really conducive to the desired results. We often talk about 'theories' and 'accounts of' or 'approaches to' conditionals as if the content of these terms were self-evident, but I suspect that it is often far from obvious what the purported explanations are supposed to be aiming at. My qualms are probably ingenuous, but I sorely miss a reflection on the default methods of philosophical inquiry into conditionals in the literature. Given the abundance of the latter, this laxity is more than a little surprising.

I can't promise you much by way of clear-cut arguments and neat conclusions. In particular, I am not going to put forward any 'alternative semantics' in this dissertation. What I will try to do is throw some light on the origins of the philosophical problem of conditionals and how they shaped the subsequent developments. I will present a historical outline of those philosophical positions on the topic that have dominated the literature since Frege, focusing on a few especially resilient puzzles along the way. I will subsequently argue that one of the reasons for our failure in tackling them may reside in our theoretical mindset—we are trying too hard to emulate Frege's feat with the material conditional for much more complex natural language phenomena, which for all we know may not be amenable to such treatment. I will claim that in natural language analysis, purely formal methods have to come to an end at some point, lest they should run the risk of sliding into triviality.

In the rest of this introductory chapter, I will very briefly acquaint you with the rudimentary distinctions and concepts ubiquitous in the field. I will then go on to present the origin of the problem of conditionals in

analytic philosophy—Frege's truth-functional conditional stroke and the thesis (to which, as we'll see, he didn't subscribe) that it is equivalent to the natural language 'if'.

Given the many shortcomings of this position, the first part of the dissertation, 'The Theories' (Chapters 2-5) will explore the most relevant theoretical alternatives. We shall review probability approaches to indicatives in Chapter 2. Cotenability theories of subjunctives are the subject of Chapter 3. The most popular theory of (mainly subjunctive) conditionals, the Stalnaker-Lewis similarity theory, will occupy us in Chapter 4. In Chapter 5, we will have a look at some theoretical possibilities that have arisen mainly in reply to the success of the similarity theory: Lycan's and Gauker's theories of conditionals, Kratzer's and Veltman's premise semantics, as well as the causal theories of subjunctives.

In the second part of the dissertation, named 'Conditionals in natural language', I try to confront my disappointment with the actual yield of this avalanche of theorising. As I have already suggested, I will argue that much of the defects of the current research programme are down to inadequate methodological assumptions. Chapter 6 is entirely devoted to indicatives, and strives to show that sometimes a simple account of meaning can accomplish more than a sophisticated formal semantics. Chapter 7 is more speculative in nature and tries to supply the missing link between indicatives and subjunctives, while bringing out the latter's role in our intellectual lives.

1.1. Basic notions

1.1.1. Conditional conjunctions

'If you're reading this, the chances are that you're going to be in Santiago de Compostela in a few weeks.' There is perhaps no natural language expression that plays such a privileged role in our inferences as the conditional conjunction[2] "if", occasionally followed by "then" in the main clause. In logic and linguistics, the subordinate clause is known as the *antecedent*, or sometimes as *protasis*, and the main clause as the *consequent* or *apodosis*.

[2]In the entire work we have to bear in mind the distinction between two meanings of the word 'conjunction'. The first is the meaning it has in general linguistics, in which it refers to any operator whose task is to create compound sentences out of simple ones. The second is the meaning it has in logic, where it stands for the truth-functional sentential connective '$\wedge$'. Whenever there is danger of confusion, I will use 'logical conjunction' for the second case.

However, we will not be concerned here exclusively with this particular linguistic expression, but rather with the sort of linguistic contribution it is designed to make. There are many other ways of putting very much the same message across in English, for instance by means of conjunctions such as "unless" or "whenever". Alternatively, you can sometimes use inversion, as in

(1) Had I known she would be here, I might have stayed home reading the latest EU regulation instead.

and

(2) Should you have any questions about that contraption, don't ask me—I have no clue.

Finally, sometimes simple stacking of sentences is all that you need:

(3) We go in there, we don't come out.

(4) Comes love, nothing can be done.

(5) No Waterloo, no Whig history.

(6) Brug open, motor uit![3]

Though the subtle differences between these options can undoubtedly teach us plenty, we will be mostly looking at what they have in common rather than at what makes them different. The latter task would most likely be better handled by linguists, not philosophers of language (although the distinction between them, if there ever was a clear one to begin with, is getting blurred by the day).[4]

1.1.2. Indicatives vs. subjunctives

It should be clear from the above examples that conditionals admit of many different kinds of propositions as arguments. They can vary in tense

[3]The examples are, respectively, from a scout's report before the Little Bighorn battle, a classic jazz song, (Ferguson, 1999), p.9, and finally, an exhortation often found on Dutch moveable bridges across navigable canals:'Bridge open, motor off!'

[4]For a view dissenting from this generous stance on conditionals, compare (Lycan, 2001), p.3, where examples analogous to (1)-(6) are considered non-conditional in nature. That seems to be chiefly down to the prominence Lycan awards to syntax in the development of his theory. There is no doubt that these examples differ in their syntax from typical if-then sentences; however, I think conditionality should be regarded as a matter of meaning rather than one of syntax. Anyway, it is the semantic dimension I am after here. For a similar conclusion, see (Rescher, 2007), pp.3-4.

and modal scope, and they can be to various degrees elliptical. Moreover, examples (2) and (6) show that the illocutionary force of the consequent does not even have to be assertive, but can be directive instead. We can also easily come up with examples of conditional promises, excuses and pronouncements.

If we restrict ourselves to purely assertive conditionals, we will find that the literature has traditionally distinguished between two broad classes of them: *subjunctives* and *indicatives*. An example of a subjunctive conditional would be

(7) If you had studied something useful, you would not have to become a PhD student.

And here goes an example of an indicative conditional:

(8) If the Neanderthals could talk, we will never be able to reconstruct their language.

Every respectable text on conditionals seems to begin with a lament on how inadequate this terminology is. This would be understandable if most theorists rejected the distinction that underlies it, yet the opposite is true. Anyway, I for one am not going to join the choir—I think that the distinction is crisp enough to begin with and the chosen terms actually quite felicitous. On an intuitive level, the difference in meaning between indicatives and subjunctives appears to be quite straightforward. Tentatively, the antecedent of an indicative presents a live epistemic possibility and the assertion made by the consequent is subject to this possibility turning out to be actual. On the other hand, the antecedent of a subjunctive typically introduces a hypothesis that is accepted as false and the consequent makes an assertion about a non-actual course of events. That's why subjunctives are often simply called *counterfactuals*.

Now, I agree that these are mere pre-theoretical hunches that have to be confirmed, elaborated on and eventually explained by serious theory, but they are strong enough to provide a good starting point for inquiry into conditionals. Moreover, the distinction between 'indicatives' and 'subjunctives', as characterised above, is very neat across many natural languages and often, though not always, it really matches the grammatical moods traditional grammarians refer to when employing these terms. For instance, Bennett maintains that the English subjunctive has little to do with 'subjunctive conditionals', but this has been put into question by Gauker.[5] In Spanish, indicatives do use the 'indicativo' mood and subjunctives the

[5]See (Bennett, 2003), p.11 and (Gauker, 2005), pp.226-240. Bennett, in my view,

past 'subjuntivo' mood, while other languages employ all kinds of different devices to mark the distinction. However, this is largely irrelevant for our purposes—one can argue that we are warranted in introducing new terminology if (i) it is useful, (ii) it doesn't give rise to confusions and (iii) is readily applicable to a wide array of intended phenomena. I wish all our terminology complied with these desiderata as much as the terms 'indicative'-'subjunctive' do.

To represent an indicative conditional with antecedent A and consequent C formally, I shall be using the simple arrow:

(9) $A \rightarrow C$

For subjunctives, I shall opt for the special boxarrow symbol introduced by D. Lewis:

(10) $A \mathrel{\Box\!\!\rightarrow} C$

Finally, I'll use Stalnaker's corner whenever no distinction is made between the two types of conditionals:

(11) $A > C$

For a long time, many theorists believed that the choice of the subjunctive form was no more than a way for the speaker to signal her lack of belief in the antecedent and that except for this minor detail, both varieties of conditionals were evaluated following the same procedure. This changed when E. Adams put the following pair on the table:[6]

(12) If Oswald hadn't shot Kennedy in Dallas, then no one else would have.

(13) If Oswald didn't shoot Kennedy in Dallas, then no one else did.

If the Warren Commission is to be trusted, then it is reasonable to believe that it was Oswald who killed Kennedy and that he acted alone. Someone who has those beliefs may, though, very well end up rejecting (13) and accepting (12). This suggests that a lack of credence in the antecedent is not the only feature that sets subjunctives apart from indicatives, yet the search for a more comprehensive explanation of the difference has proved surprisingly difficult. Before plunging into it, though, a few words about the material conditional are in order.

---

erroneously focuses on the main clause. Gauker also attacks the arguments against the syntactic distinction between subjunctive and indicative mood in English marshalled in (Dudman, 1988). Dudman's arguments do not translate into other languages (say, Slavic ones), which do seem to mark this very semantic difference in a systematic way. This suggests that syntax is not always a reliable guide to semantics.

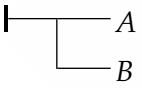
[6](Adams, 1970), p.90

1.2. The material conditional

1.2.1. Forefathers

Frege

The material conditional made its entry into logic in 1879, when Frege defined the meaning of the conditional stroke in his *Begriffsschrift* as follows:

> If A and B stand for contents that can become judgments (§2), there are the following four possibilities:
>
> (1) A is affirmed and B is affirmed;
> (2) A is affirmed and B is denied;
> (3) A is denied and B is affirmed;
> (4) A is denied and B is denied.
>
> Now
>
> 
>
>
> stands for the judgment that *the third of these possibilities doesn't take place, but one of the three others does.*
>
> (Frege, 1967), pp.13-14

The affirmability of Frege's conditional stroke thus depends functionally on the affirmability or deniability of the contents in its antecedent and consequent—its semantic contribution doesn't go beyond banning the possibility of affirming the antecedent and denying the consequent.

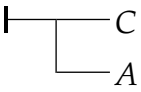
Frege was explicit about the fact that he wasn't putting this technical device forward as an accurate counterpart of the common conditional. He claimed that the conditional stroke could only be translated by means of 'if' on the condition that it was asserted without clear commitment to either the antecedent or the consequent. And even in such cases, Frege felt that 'if' conveyed an idea of a causal link between the conjoined contents that was absent in his conditional stroke.[7]

Unlike most of his followers, Frege didn't consider this divergence to be a reason for worry. As one of the pioneers of formal languages, he seems to have had a clearer idea of their limitations than many of those who later came to take them for granted. The job of the system (at least at that

[7](Frege, 1967), p.14

stage of its development) was to allow precise rendering of mathematical reasoning, and Frege repeatedly suggested that this objective could not be achieved without pruning natural language of many of the features irrelevant for this particular project. A formal language is as specialised a tool as a microscope—its very precision precludes its use for everyday purposes. If mathematicians managed to cram their axioms and theorems into the mould provided by the artificial conditional stroke, then nothing more could be asked of it.[8]

At a later stage, Frege abandoned the somewhat pragmatic-sounding talk of affirmation and denial (*Bejahung* and *Verneinung*) for the talk about truth-values as referents of the thoughts expressed by sentences combined by the conditional stroke. In the new idiom, a conditional such as

(14) 

refers to the Truth (or simply, is true) whenever *C* refers to the Truth or *A* refers to the Falsity. This is more or less the interpretation under which Frege's conditional stroke, and, for that matter, the whole of his logic, was adopted by posterity.[9]

On the other hand, and more importantly, Frege doesn't seem to have undergone any radical change of mind about the relation between logic and ordinary language ('die Sprache des Lebens', as he called it), as the following passage attests. It comes from his *Gedankengefüge*, finished a mere two years before his death, and it deals with the question whether the binary operation on judgeable contents corresponding to the conditional stroke can be adequately rendered by 'if':

> But here, indeed, doubts may arise. It may perhaps be maintained that this does not square with linguistic usage. In reply, it must once again be emphasized that science has to be allowed its own terminology, that it cannot always bow to ordinary language. Just here I see the greatest difficulty for philosophy: the instrument it finds available for its work, namely ordinary language, is little suited to the purpose, its formation having been governed by requirements wholly different from those of philosophy. So also logic is first of all obliged to fashion a useable instrument from those already to hand. And for this purpose it initially finds but little in the way of useable instruments available.
>
> (Frege, 1963), p.11

[8] See (Frege, 1967), especially pp.6 and 12.
[9] See (Russell and Whitehead, 1964), p.7.

To sum up, Frege's invention of the conditional stroke was never meant as an attempt at a precise analysis of natural language—there was no questioning of its artificial nature.

Russell

After proving Frege's attempt to justify logicism to be inconsistent, Russell took it from where Frege left off. He embraced Frege's use of the conditional stroke in logic, but found his two-dimensional formulae rather impractical and opted for Peano's notation instead. Eventually it was the notation and terminology adopted in the *Principia Mathematica* which set the standard in the literature.

In the new notation, the conditional stroke is replaced by the one-dimensional symbol '$\supset$', usually called *horseshoe* in the literature. It comes originally from the letter C, placed by Peano between two propositions in order to affirm that the first is a *consequence* of the second. An inverted C stood for the inverse of this relation and gradually evolved into the horseshoe as we know it today.[10] Frege's (14) now becomes the much more economical and perspicuous

(15) $A \supset C$.

The truth-function expressed by these conditionals, which is of course identical to that of Frege's conditional stroke, was dubbed 'material implication' by Russell. That's why I shall henceforth call the substitution instances of (15), in accordance with a widespread usage, 'material conditionals'. The term 'implication' was warranted, according to Russell, because $A \supset C$ amounted to the most straightforward ground upon which C could be inferred from A.[11] And Russell called it 'material', so as to distinguish it from the 'formal implication', which he defined as a relation between two propositional functions, as stated by the formula

(16) $\forall x(\Psi(x) \supset \Phi(x))$

Russell claimed that our uneasiness about the notion of material implication were chiefly due to conflating it with its formal counterpart. If we obtain the statement of a material implication without deriving it from one of a formal implication, it is usually because we have already proved the falsity of the antecedent or the truth of the consequent.[12] Yet in either of

[10](Sanford, 1989), p.51
[11](Russell, 2010), p.34
[12](Russell and Whitehead, 1964), p.20-1

these two cases, the material implication is rather useless for the purposes of inference—either we had already proved what we were after, or we can never use the implication in a modus ponens, given that the minor premise is ruled out.

Observe that if a material implication is derived in a sound system of pure logic, then it must be logically true, which in turn means that C follows logically from A. This distinction presupposes a semantic notion of consequence and as such could not be made by Russell, but it does explain the rationale behind his views on the conditional connective he used. The only intuitively problematic cases in *Principia Mathematica* are therefore instances of the Principle of Explosion and its converse, but these are easier to accept in the context. It is only if the material implication is wrenched from its natural habitat, i.e. a classical deductive calculus, that the horseshoe displays paradoxical behaviour.

1.2.2. Paradoxes

Under *paradoxes of material implication* we understand a large group of valid patterns of inference and logical truths involving the material conditional, which sound unacceptable if the horseshoe is replaced by the ordinary 'if-then'. The most famous ones are

(CAC) $C \therefore A \supset C$

and

(NAAC) $\neg A \therefore A \supset C$

If the first inference pattern is translated into natural language with the help of an 'if', it will justify the inference from any proposition C to a conditional with C in the consequent and any proposition A whatsoever in the antecedent. This seems to validate, for example, the following piece of reasoning:

(17) Tomorrow is Friday. Therefore, even if I've mixed up the dates, tomorrow is Friday.

If, on the other hand, we do the same with the NAAC pattern, we will be compelled to accept any conditional introducing a hypothesis incompatible with an accepted proposition:

(18) The stores do not open tomorrow. Therefore, if the stores open tomorrow, they won't have any customers.

And there is more. The following formula is a logical truth:

(19) $(A \supset B) \vee (B \supset A)$

So for any two formulae, either one implies the other or vice versa. Again, this can be warranted in a classical deductive calculus, where the principle of bivalence guarantees that any proposition will either be inferrable (if true) or unavailable as the minor premise in a modus ponens (if false). However, a natural language translation can sound very bizarre indeed:

(20) Either if I get extra holiday this year, I'll spend an entire week reading Hegel, or if I spend an entire week reading Hegel, I'll get extra holiday this year.

Other important inference patterns valid on the material interpretation that nonetheless have suspicious substitution instances when rendered by natural language ifs (whether indicative or subjunctive) include Antecedent Strengthening (AS), Hypothetical Syllogism (HS) and Contraposition (CP).[13]

There is not much evidence that Russell disagreed with Frege about the technical nature of his notion of implication. In particular, he doesn't seem to have intended his material conditional as an analysis of the natural language 'if':[14]

> And although, while relations are still regarded with the awe caused by unfamiliarity, it is natural to doubt whether any such relation as implication is to be found, yet [...] there must be a relation holding between nothing except propositions, and holding between any two propositions of which either the first is false or the second true. Of the various equivalent relations satisfying these conditions, one is to be called implication, and if such a notion seems unfamiliar, that does not suffice to prove that it is illusory.
>
> (Russell, 2010), p.35

Arguably, if the material implication were considered by Russell to be at the semantic core of our everyday ifs, he could not have deemed it that unfamiliar to a common reader, let alone to a mathematically sophisticated

[13]Counterexamples to these and other inference patterns have become something of a stodgy fare in the literature on conditionals, so I feel free to refer you to some of the classics for details: (Stalnaker, 1968), pp.48-49; (Lewis, 1973a), pp.31-36; (Sanford, 1989), pp.95-97 and 108-111; (Edgington, 2007), pp. 143-145; (Woods, 1997), p.25; (Bennett, 2003), pp.138-145 and 159-168; and many more.

[14]But see (Sanford, 1989), p.65.

one, after all. This of course does not mean that this truth-function cannot appropriately be rendered by means of 'if-then' in the context of a classical deductive calculus, which of course Russell explicitly admits—this may well be one of the tasks to which ifs are particularly well-suited. However, this is still a far cry from an argument for semantic conflation of indicative and material conditionals.

The question about their relationship was bound to be asked at some point, though, especially with the boom of logically-minded analyses of natural language that began with logical positivists. Natural indicative conditionals validate, by and large, inferential connections (either essential or merely accidental ones) between propositions, wherefore they are fitted for use in Modus Ponens. Yet so do material conditionals. If uncertainty about the truth value of the component propositions, alluded to by Frege, were all there was to the difference between the two, then maybe the perceived awkwardness of some uses of the material conditional could be explained by pragmatic, rather than semantic considerations. This is the idea that inspired Grice's theory of implicature.

1.3. Identity Thesis

1.3.1. Conversational Implicature

One of the philosophically most fruitful defences of the identity thesis (which identifies ordinary 'if' with the material conditional) was carried out by H. P. Grice in his William James Lectures in the 1960s. Despite the enormous success of his theory of conversational implicature, which has spread well beyond the narrow confines of philosophy of language, many tend to forget that Grice's foremost reason for advancing it was to provide a good explanation of the relation between logical devices and their alleged natural language counterparts. For Grice, such an explanation had to fulfil two rather conservative criteria: it should not put into question the validity of classical first-order logic as the logic of the ordinary English, and it should be wary of cluttering our semantics with a multiplicity of new meanings. This section will be devoted to an outline of Grice's position, as well as to the criticism it has stirred among dissenting theorists.

Modified Occam's razor

I will not start our exposition head-on with an analysis of the celebrated notion of implicature, as I believe that in order to approach it correctly, we

have to set it in the broader context of Grice's general semantic project. Grice made use of this notion in order to explain why an expression could contribute to the meaning of an utterance in several different ways, without attributing multiple conventional meanings to it. The seemingly anomalous behaviour of an expression should thus be explained by conversational interplay between a speaker and her audience. The ultimate goal of this strategy is to maintain linguistic meanings as lean as possible and put their apparent proliferation down to pragmatic phenomena.

Grice introduces a heuristic rule of thumb called Modified Occam's Razor (MOR), which underlies his project of semantic analysis. According to the MOR, the entities that must not be multiplied beyond necessity are meanings. Or, in an alternative formulation put forward by Grice himself, no derivative meaning should be attributed to a piece of vocabulary if it is possible for an audience to retrieve this meaning without possessing additional semantic knowledge.[15] Grice adds that if we accept the further supposition that it is easier to contextually augment rather than diminish the logical force conventionally conveyed by a sentence, then it follows that our semantic analysis should not only render our vocabulary as univocal as possible, but also as logically weak as possible.[16]

Conversational implicature

Implicature, according to Grice, is a device used by a speaker to communicate information that goes beyond the strictly truth-conditionally evaluable content conveyed by an utterance.[17] Hence, the implicatum is not a logical consequence of the proposition expressed by the speaker. We speak about *conventional implicature* when the audience retrieves this information by dint of the linguistically codified meaning of the vocabulary used by the speaker, which nonetheless does not contribute to the truth conditions of the utterance. By contrast, if this additional information is obtained through a pragmatic interpretive inference guided by the presumption of rationality of the speaker, we are dealing with a *conversational implicature*. We will have more to say about the former when discussing the theories by Jackson and Strawson; at this point we will focus on the conversational

[15](Grice, 1978), p.47-48

[16]Grice admits that he can't justify this additional principle, but he finds it very appealing intuitively. We will see that this latter premise is tacitly present in Grice's argumentation in favour of the Identity Thesis, but we have to bear in mind that it is independent from the main clause of the MOR. The additional requirement does not automatically inherit the MOR's theoretical plausibility.

[17](Grice, 1975), pp.24-25

implicature, which lies at the core of Grice's take on the counterintuitive consequences of the identity theory.

Suppose you have the following conversation with a friend:

You: I'm going to do some shopping.
Friend: It's May 1.

Although quite trivial, this exchange would be quite unintelligible should we pay attention only to the explicit, truth-conditionally evaluable content involved in it. So as to understand your friend's contribution to the exchange, you have attribute to him belief in the following proposition:

(21) On May 1 the shops are closed.

This content does not appear explicitly anywhere in the above conversation, but its availability to both you and your friend is crucial to the success of the exchange. Neither is it a logical consequence of any other explicit proposition involved in the conversation and it might change, or even be unavailable, were the same explicit content conveyed in special circumstances of utterance. To put it in a more straightforward way, there are propositional contents that the audience must retrieve in order to preserve the presumption that the speaker is trying to make her utterances as conversationally helpful as possible, or, which is the same, the presumption that she follows the *Cooperation Principle*, as set forth by Grice.[18] These contents receive the name of conversational implicatures.

Grice distinguishes between generalised and particularised conversational implicature.[19] The former is invariably triggered in typical occasions of utterance of the proposition in question, while the latter requires a specific context to arise. A paradigmatic example of a generalised conversational implicature is supplied by our common use of the existential quantifier:

(22) Some of the students took part in the meeting.

typically implicates

(23) Not all the students took part in the meeting.

but we certainly would not want to say that (22) entails (23). The pragmatic inference that leads to the conclusion (23) is based on the reasonable

[18](Grice, 1975), p.26
[19](Grice, 1975), pp.37-38

assumption that if the speaker were committed to a logically stronger proposition than (22), such as 'All the students took part in the meeting', it would be appropriate for him to affirm the latter proposition instead of the former. This is the general strategy used by Grice in order to explain the behaviour of logical vocabulary in natural language.

Disjunction, conditionals and conversational implicature

Grice's defence of the identity thesis for conditionals and disjunctions is exposed respectively in (Grice, 1967) and (Grice, 1978), and follows more or less the same lines. In both cases he urges us to consider the arguments of a 'strong theorist', who claims that the meaning of the natural language counterparts of our logical devices consists of a truth-functional core and a further indication (called 'the indirection condition' in the case of conditionals) to the effect that 'there are non-truth-functional grounds for affirming them'. In this way the 'strong theorist' can explain the intuitive invalidity of inferential patterns like (CAC) or (NAAC).[20]

In both cases the only ground for asserting the conditional would reside in our previous knowledge of the truth value of either the antecedent or the consequent. While this is sufficient for the truth of a material conditional, it cannot warrant an assertion of the corresponding indicative conditional, according to the strong theorist. The same is true for disjunction—we would most certainly object to an assertion of 'A or B', should it be motivated solely by the speaker's belief that A is true.

Grice points out that the MOR requires us not to regard this non-truth-functional extra as a part of the linguistic meaning of our logical vocabulary, if we are capable of explaining its emergence as a case of generalised conversational implicature. He then tries to come up with such an explanation.

A speaker who asserts '$A \supset C$' exclusively on the basis of his belief in the truth of C violates the conversational maxim 'Make your assertion as logically strong as possible'. The assertion of '$A \supset C$' or '$A \vee C$' under these circumstances would be, to use a phrase of Stalnaker's, 'pointless, hence misleading'.[21] This is why an audience with no reason to suspect that the speaker is not following the Cooperation Principle typically retrieves the implicature about non-truth-functional grounds for the speaker's assertion.

There is more evidence to the effect that the 'indirection condition' is

[20] See page 10.

[21] (Stalnaker, 1975), p.147

actually a conversational implicature—it possesses some of its defining traits. First, it is quite non-detachable from the truth-conditionally evaluable content of the utterance, as it is trigged by the assertion of any formula classically equivalent to the material conditional in question. And second, there are certain circumstances of utterance under which the implicated content is cancelled, whether by means of speaker's explicit instruction, or contextually.

Conversational implicature in D. Lewis

Grice was not the only theorist to resort to the theory of conversational implicature to explain the quirks of indicative conditionals in natural language. David Lewis, who had applied possible-world semantics to past subjunctive conditionals, adopted a variety of Grice's theory for indicatives in order to explain the intuitions at the base of E.W.Adams's work without sacrificing the truth-functional account of indicative conditionals.[22] As we will presently see, Adams's probabilistic conditionals verify

(24) $p(A \rightarrow C) = p(C/A)$

which in general differs from $p(A \supset C)$.

Lewis felt compelled by his triviality results to reject the thesis that probabilities of indicative conditionals are conditional probabilities. Nevertheless, he was also interested in showing that there was an intimate connection between the acceptability of a conditional and the conditional probability of its consequent with respect to its antecedent. The challenge, as with Grice, consisted in explaining this connection while keeping the linguistic meaning of the indicative identical to that of the material conditional. In order to defend the identity thesis, Lewis adduced the impressive success of classical logic and the fact that embeddability of probabilistic conditionals in truth-functional formulae would have to be severely constrained in order to avoid triviality. The upshot was, according to Lewis, that natural language indicative conditionals could not be Adams's probabilistic conditionals.

Lewis, as well as Grice, suggested that a speaker should not assert a conditional only on the strength of her knowledge of the falsity of the antecedent. If in addition the conditional probability of the negation of the consequent over the antecedent is high, the conditional will most certainly lack in assertibility. The formula that expresses the total loss of assertibility will then be

[22](Lewis, 1976), pp.137-139. See Chapter 4 for details of Lewis's take on subjunctives.

(25) $p(\neg A) \times (p(\neg C/A))$

hence the formula expressing the resulting assertibility of a material conditional would be:

(26) $As(A \supset C) = p(A \supset C) - p(\neg A) \times p(\neg C/A) = p(C/A)$

Lewis went on to say that the same reflection was applicable to any other logical constants capable of expressing the same truth function as the material conditional.

Criticism of Grice's and Lewis's account

The elegance of Grice's theory has been unanimously acknowledged and only reticently have many theorists rejected it as an adequate explanation of the differences between mathematical and natural conditionals. First, I will expound the criticism other authors have levelled at Grice, and afterwards I will present some general considerations that make me share their opinion.

Frank Jackson offers examples of acceptable conditionals with very improbable antecedents and others with very probable consequents.[23] An instance of the former type would be

(27) If the sun goes out of existence in ten minutes' time, the earth will be plunged into darkness in about eighteen minutes.

and of the second type

(28) Even if I vote against the candidate X, he will still win.

On Grice's account, these conditionals should strike us as mischievous, as it would be much more informative to simply state their antecedents and consequents, respectively. The fact that they don't suggests that there is more to assertibility of conditionals than their logical force and probability of the embedded propositions. We could go even one step beyond Jackson's examples and consider a conditional that would be perfectly acceptable in an arithmetic reductio, despite having an impossible antecedent.

(29) If $m + n \neq n + m$, then $(0^m)^n \neq (0^n)^m$

[23](Jackson, 1987), pp.20-21.

What is more, Jackson calls into question even the crucial premise of any Gricean treatment of conditionals—it seems that a speaker may legitimately opt to affirm a logically weaker proposition in retribution for a considerable gain in robustness, that is, the ability of the affirmation to withstand evidence to its contrary. Bennett considers Jackson's objections to Grice's and Lewis's theory as conclusive.[24]

Dorothy Edgington focuses on the distinction between 'learning a proposition is true' and 'finding it fit for assertion', which is at the heart of Grice's and Lewis's proposals.[25] Her conclusion is that there is no evidence of competent language speakers recognising this tension in their everyday use of conditionals. Even though her objection is expressed on a fairly reduced space, we can develop the intuition behind it invoking Grice's own dictum as to the distinction between 'what is said' as opposed to 'what is implicated' by the utterance:

> If nonconventional implicature is built on what is said, if what is said is closely related to the conventional force of the words used, and if the presence of the implicature depends on the intentions of the speaker, or at least on his assumptions, with regard to the possibility or nature of the implicature being worked out, then it would appear that the speaker must (in some sense or other of the word 'know') know what is the conventional force of the words he is using.
>
> (Grice, 1978), p.49

Here Grice seems to be suggesting that the proposition expressed by an utterance by virtue of the linguistic meaning of the chosen vocabulary must be consciously accessible to the speaker in order to trigger any implicatures. A similar position is advocated by the contextualist approach to the line of divide between philosophical semantics and pragmatics.[26] Undoubtedly, it would be very strange if competent speakers were so wrong about the use of the conjunction 'if' that they would not be able to identify the inferences justified only on the strength of its conventional meaning, without resorting to any contextual criterion of assertibility.

As for my own criticism, I would like to speculate that the reason Grice's take on conditionals has not been as successful historically as its ingenuity might have us expect was its failure to comply, if not with the letter, then at least with the spirit of the MOR.

---

[24](Bennett, 2003), p.33

[25](Edgington, 1986), pp.9-10

[26]See for instance (Blakemore, 2002), (Recanati, 2004) or (Wilson and Sperber, 2004). However, this objection had already been raised by Mackie: (Mackie, 1973), p.80.

We have seen that it is a reasonable requirement for our semantics not to attribute more derivative meanings to a piece of vocabulary than those strictly required to explain its functioning in a language. However, Grice has not made any attempt to justify the additional and independent requirement to the effect that the attributed conventional meaning should be as logically weak as possible. And it seems that the loss of generality incurred by Grice's theory of indicative conditionals is too high to warrant this additional assumption in case of indicative conditionals. The fact is that Grice offers at least three further independent pragmatic explanations in addition to the one we have discussed so far in order to account for yet other anomalies entailed by his defence of the identity theory based on conversational implicature.[27]

For instance, he is forced to face the generally acknowledged intuition that negations of conditionals are sometimes understood as conditional negations of their consequents—an intuition that cannot be dealt with by appealing to the maxim 'make your assertion as logically strong as possible'. He solves the problem by introducing an ad-hoc interpretive mechanism which conveniently alters the syntactic order of the words in uttered sentences. What is more, he admits his incapacity of explaining the failure of the law of contraposition in examples where conditional probability is involved. However, a probabilistic account of indicative conditionals can cope with all these alleged paradoxes in one fell swoop.

Theorists working under what we will call 'the Ramseyan paradigm' have interpreted these problems as evidence that a unified semantic account of indicative conditionals would be preferable to Grice's theory, even if it were forced to attribute more logical force to conditionals. However, Grice's work offers yet another possible way of escape for identity theorists, by means of his theory of *conventional* implicature. This path was be chosen by F. Jackson and P. Strawson.

1.3.2. Conventional Implicature

Once the attempts to explain away the paradoxes of material conditional by means of conversational implicatures failed, it seemed unavoidable to conclude that the inferential behaviour of conditionals in contexts characterised by imperfect certainty must somehow be codified in their linguistic meaning. This realisation, however, need not imply an abandonment of the identity thesis, although it must be slightly modified. Conventional implicatures are an identity theorist's last stand—they allow him to preserve

[27] See (Grice, 1967), pp.83-85.

truth-functionality of indicative conditionals, while he can also maintain that the expectation that the conditional probability of the consequent with respect to the antecedent be high arises by default and not by virtue of any pragmatic interpretive mechanism. Whereas F. Jackson draws on Grice's theory of conventional implicature explicitly, I believe that P. Strawson's theory of conditionals presupposes something very like that.

Conventional implicature in Grice

Unfortunately, Grice's loci classici for the study of the subject of conventional implicature are very spare and fairly succinct. In his (Grice, 1975), Grice devotes to this linguistic phenomenon only a few brief remarks:

> In some cases the conventional meaning of the words used will determine what is implicated, besides helping to determine what is said. If I say (smugly), *He is an Englishman; he is, therefore, brave* I have certainly committed myself, by virtue of the meaning of my words, to its being the case that his being brave is a consequence of (follows from) his being an Englishman. But while I have said that he is an Englishman, and said that he is brave, I do not want to say that I have *said* (in the favoured sense) that it follows from his being an Englishman that he is brave, though I have certainly indicated, and so implicated, that this is so. I do not want to say that my utterance of this sentence would be, *strictly speaking*, false should the consequence in question fail to hold. So *some* implicatures are conventional, unlike the one with which I introduced this discussion of implicature.
>
> (Grice, 1975), p.25

To sum up the passage, the speaker commits himself to the content implicated conventionally by the linguistic meaning of his words, but this content does not enter into the truth-conditional evaluation of the utterance. The truth conditions of an utterance determine 'what is said', not 'what is implicated'.

Other classical formulation of the theory can be found in (Grice, 1961). In this article Grice compares the properties of two different types of content suggested, as opposed to entailed, by the linguistic meaning of the words used in a sentence.

(30) Smith has left off beating his wife.

The proposition suggested by (30),

(30') Smith used to beat his wife.

has the status of a *presupposition*, that is, its truth is a necessary condition for (30) to have a truth value. The situation is different with

(31) She was poor but honest.

The content suggested by the contrastive conjunction 'but', according to Grice, is the existence of some kind of opposition between poverty and honesty. This content, he argues, is independent from the truth-conditional assessment of the sentence (31), which is equivalent to

(31') She was poor and honest.

It is of crucial importance, therefore, not to conflate two semantic categories in Grice: on the one hand, the conventional meaning of the expressions used in an utterance, and on the other hand, 'what is said', that is, the propositional content that determines the truth conditions of an utterance.

Now that we have sketched out Grice's original position, we should inquire into its connection with the above-discussed Modified Occam's Razor. Grice is fairly unambiguous on this point: the MOR mediates between the contents retrieved by pragmatic interpretive inferences and linguistically codified meanings. It enjoins us to attribute linguistic meanings as univocal and (according to a further assumption) logically weak as possible to our vocabulary. It follows that the MOR is of little help in distinguishing between the ingredients of linguistic meaning that contribute to the truth conditions of an utterance and those that make up the conventional implicatures conveyed within the speech act. To draw this line is an extremely elusive task, which, according to Grice's view, should be guided fundamentally by the intuitions of competent language users.[28] It seems that if competent language practitioners consistently attribute the same truth value to two sentences, one of which uses words that are somehow semantically stronger than those of the second one, this additional content conveyed by the former sentence should be explained by an appeal to a conventional implicature triggered by the words in question. Language users will typically perceive some contextual violence if the weaker expression is substituted for the stronger one in inadequate contexts, but they will not be prone to deny the truth of the proposition expressed under these circumstances, on condition that the weaker proposition was originally considered true by them. In paradigmatic cases, the audience will object to the tone of the utterance, but not against its explicit content.

Consider the following utterance:

[28] For a good review of this question, see (Bach, 1999).

(32) Not even a student can afford such a holiday.

We will tend to regard (32) as true if the holiday costs exceed the income of a common student, but even then we will not be happy with the content suggested by the utterance of this sentence, which is that students enjoy above-average incomes.

If this rendering of Grice's account is viable, we might think that the relation between the conventional implicatum and what is said is the same as that between the conversational implicatum and linguistic meaning—the task of semantic analysis would consist in making 'what is said' as logically weak as our linguistic intuition allow.

The theory of conventional implicature can be and was applied to the analysis of conditionals. They would present two layers of conventional meaning: one that specifies their truth conditions (i.e., if the identity thesis is endorsed, those of the corresponding material conditional) and an additional rule that would specify the adequate contexts for their utterance (thus making them *prima facie* logically stronger). Once we have dismissed Grice's original theory based on conversational implicature, this is as close as we can get to the full-fledged identity thesis.

Conventional implicatum as inferential connection

We can interpret Strawson's criticism of Grice's theory of indicative conditionals as making an implicit use of a concept akin to that of conventional implicature.[29] Strawson himself strove to defend the 'strong theorist's' position, without being familiar with the details of Grice's proposal.

His theory draws on the close relationship between the sentential operators 'if' and 'so'. The main argument in favour of introducing the strong condition of existence of non-truth-functional grounds for assertion in the conventional meaning of sentential structures built with an if-operator is that Grice's analysis would make it impossible for a language to include an operator with a meaning analogous to 'so' that could relate non-asserted propositions. However, the linguistic meaning of 'so' resides in the very indication of a relation of dependence between the propositions it connects, and to which the speaker commits herself. A semantic analysis guided by the MOR doesn't seem to be able to uncover any more fundamental layer of its meaning (although its truth conditions are just those of the logical conjunction).

Bennett criticises Strawson's account, pointing out that there are many acceptable conditionals where this strong connection between the an-

[29](Strawson)

tecedent and the consequent is absent.[30] Nevertheless, Strawson explicitly states that this conditional link may be as weak as one can imagine and often is completely context-dependent. According to Strawson, 'so' functions as an inference marker even in compound sentences such as

(33) Melbourne is in Australia, so the sea is salt.

where it is exclusively the context which determines the acceptability of the inference (think, say, of a multiple-choice test with unique solutions).

However, even if Bennett's critique should be viewed as misguided, attention must be drawn to the purely negative character of Strawson's theory—he does not develop an independent proposal, he only purports to reply to Grice's work. And the manner in which he does so is controversial: precisely because 'so' always connects propositions to which the speaker commits herself, if it has any additional meaning compared to the logical conjunction this cannot be explained by conversational implicatures. However, since conditionals typically feature propositions whose truth value is left undetermined, this kind of analysis remains viable. And, unless we reject the MOR, if it is viable, it should be preferred to others.

Strawson's proposal did not attract many followers[31], but we should underline the strategy that lies at its foundations. Structures featuring the particle 'so', appealed to by Strawson to warrant his 'consequentialist' account of if, are considered by Grice's followers as paradigmatic cases of conventional implicature . Strawson himself compares it with 'but', which suggests the presence of a contrastive, rather than consequential relation between the propositions it connects. That is why we have interpreted Strawson's position as an account of conditionals in terms of truth conditions equivalent to those of $\supset$ and a further assumption codified linguistically about the presence of a consequential relation between the antecedent and the consequent. From this point of view, Strawson's theory is much more conservative than those of the theorists who deny that conditionals have truth conditions (see next chapter) and it can be reproduced in Gricean terms.

Conventional implicatum as inferential availability

F. Jackson defended a view analogous to Strawson's, but advocating a different conventional implicatum for indicative conditionals. According to Jackson, instead of suggesting that there is an ad-hoc inferential connection

[30](Bennett, 2003), p.45.
[31]See again (Bennett, 2003), p.45.

between the antecedent and the consequent, indicative conditionals conventionally signal that they are fit for their fundamental purpose—their use in Modus Ponens.

Jackson finds it necessary to replace Grice's account based on conversational implicature for reasons we have expounded above. It is fundamental that we realise that all arguments against Grice are based on application of probability to deductive logic. Grice himself probably would have viewed accounting for reasoning under circumstances of imperfect certainty as extraneous to his project, as he sought a uniform conditional logic. Lewis's first proposal, as well as Jackson's, attempts to bridge the gap between the two logics. Lewis later acknowledged the theoretical superiority of Jackson's account and abandoned his own proposal based on conversational implicature.[32]

To explain how ordinary conditions of utterance affect the behaviour of our logical vocabulary, Jackson introduces the technical term of *robustness*.[33] A proposition p_1 is robust with respect to p_2 for a given belief system if and only if the speaker affirming it will not abandon her belief that p_1 should she learn that p_2 obtains. It follows that robustness is a relative affair, that is, the robustness of p_1 will vary with different values of p_2. Moreover, robustness is a matter of degree, as the speaker may be more or less disposed to discard p_1 on learning p_2. The third characteristic of robustness is that it is always a function of the system of beliefs of a particular speaker and not an intrinsic semantic property of the vocabulary used in assertion. It might for instance be argued that our mathematical beliefs should be maximally robust with respect to any possible evidence; however, since our rationality is finite, in practice they often fail to fulfil this desideratum.

Learning the speaker's subjective robustness of certain propositions with respect to others may be quite interesting for the audience. This information may often be retrieved by carrying out inferences guided by Grice's Cooperation Principle. With most disjunctive assertions, the speaker signals their robustness with respect to the elimination of each of the disjuncts. For instance, when asked about Pedro's whereabouts, a speaker who replies by affirming

(34) Pedro is in the kitchen or in his study.

typically commits herself to affirming

(35) Pedro is in his study.

[32](Lewis, 1986b), p.584, fn 6.
[33](Jackson, 1987), pp.22-25

should she learn from a reliable source that Pedro is not in the kitchen.

However, as this interpretation is based on a conversational implicature, it can be explicitly cancelled by the speaker, as Grice shows.[34] A speaker, playing hide-and-seek, could tell the fellow players

(34′) Pedro is in the kitchen or in his study, but I won't tell you which one it is.

By such an utterance, the speaker makes it clear that she believes the disjunction in virtue of her believing one of the disjuncts and she would therefore withdraw her credence from the former should she abandon the latter.

This conversational implicature is absent in the utterances of certain classes of disjunctions, in which one disjunct entails the other, and the speaker would therefore abandon the entire disjunction should she change her mind as to the second one. Jackson points out that in these cases we use words that signal unilateral robustness:

(36) He is in the kitchen, or anyway, somewhere in the house.

The situation with conditionals is analogous, according to Jackson. Our purpose in carrying out conditional assertions lies in entitling our audience to use them in Modus Ponens. We require that a speaker who affirms a conditional commit herself to the inference from the antecedent to its consequent. However, if the material conditional $A \supset C$ is affirmed chiefly on the grounds that the speaker believes $\neg A$ is true, detachment with A (if it despite all appearances turns out to be true) does not guarantee that the inference should be truth-preserving. This is why Jackson argues that beside determining its truth conditions, the conventional meaning of an indicative conditional includes an additional assertibility rule, by which the speaker commits herself to the validity of the corresponding Modus Ponens, should the antecedent turn out to be acceptable. In Jackson's idiom, an indicative conditional signals robustness of the corresponding material conditional with respect to the antecedent. Formally,

(37) $As(A \rightarrow C) = p(A \supset C/A) = p(C/A)$

We have shown that if conditional assertibility is construed in this manner, it will equal the conditional probability of the consequent over the antecedent, that is (as happened previously in the case of Lewis's formula), we obtain the desirable Adams's thesis on independent grounds and retaining the spirit, if not the letter, of the identity thesis.

[34](Grice, 1978), pp.44-45

We must give Jackson's theory its due—it has several advantages. One of them is that it deftly manages to avoid Lewis's triviality results[35], by denying the premise that the probability of a conditional is the conditional probability of its consequent over its antecedent. Since the truth conditions of an indicative conditional 'If A, then C' are those of the corresponding material conditional, its probability equals the probability of the disjunction $\neg A \vee C$. Another point in favour of this theory is that it reconciles the conservative tradition with the more radical approach epitomised by Ramsey and Adams, and it seems to explain the link between our logical practice and the functioning of the run-of-the-mill natural language conditionals.

Nevertheless, by many critics' lights[36], the theory also has significant drawbacks. Jackson argues in favour of the thesis that an indicative possesses the truth conditions of the corresponding material conditional, claiming that it would be discriminatory to deprive it of them, as most language structures also have them.[37] Now this is undoubtedly a highly controversial claim, but the real problem is that it makes the relation between our logic and our natural language even less clear than before. He affirms that

> [...] one who knows only (Adams) lacks nothing in the way of linguistic understanding of indicative conditionals-—after all, how many competent speakers of English know about the truth conditionas of indicative conditionals, or, indeed, about truth-conditions at all, in the philosophical sense? But this is consistent with insisting that one who does not know that $A \rightarrow B$ is true if and only if $A \supset B$ lacks something in the way of philosophical understanding.
>
> (Jackson, 1987), p.58

The lack of generality of this proposal is in my view due to its failure to attribute consistent semantic function to truth conditions over the set of expressions of our language. In a familiar model, to know the truth conditions of a sentence means to know how to constrain the set of epistemic possibilities in view of its truth conditions. In this way, the speaker and her audience can exchange information about the world.

That's why it would be extremely paradoxical to find a natural language expression, used over and over again in our daily language practice, whose truth conditions we would systematically fail to recognise. The knowledge

[35] See Section 2.3.1.

[36] See (Edgington, 1986) or (Bennett, 2003), pp.39-44.

[37] (Jackson, 1987), p.57

of the proposition expressed by such an expression would not be a matter of the man in the street, but rather require an esoteric training at philosophy faculties. What kind of role could these truth conditions play in communication? Jackson himself implicitly admits that there is none. But then it would be extremely interesting to have a look at the analysis by which a philosopher isolates the truth conditions of an expression that competent speakers use without being aware of them.

Of course, users of language may be wrong about the metalinguistic explanation of their utterances and we all carry out inappropriate speech acts at times. However, the possibility of practically all users of language being wrong about when an expression they use on a daily basis is to be considered true or false is almost a contradiction in terms. We have seen that the original Grice's theory draws heavily on competent speakers' intuition in order to identify the presence of a conventional implicature (there is no test analogous to the one with conversational implicature). This suggests that we should be at least wary of the unintuitive consequences of Jackson's theory.

1.3.3. Conclusion

It seems that the main motivation behind the defence of the identity thesis are worries about the link between natural-language reasoning and our logical calculi, which would supposedly be endangered should we abandon the material analysis of our natural-language conditionals. The general view among Gricean theorists seems to be that a change of paradigm would imply sacrificing too many of our achievements in logic and its philosophy.[38] Dissenting authors, however, maintain that many of these supposed achievements are illusory.

When discussing the alternative accounts I will attempt to show that many of these worries are unfounded and due to an erroneous conception of the task of our logic—particularly the thesis that one of the objectives of logic is to provide us with the meaning of our natural-language logical vocabulary. I believe the right thing under these circumstances would be to turn Grice on his head and start analysing the behaviour of the natural indicative conditional and then attempt to show under which circumstances material conditional would be a good approximation to it.[39] If this could be achieved, the problem would probably be resolved to the satisfaction of both parties involved in the strife.

[38] See (Lewis, 1976), pp.136-137.

[39] See Section 6.3.3.

However, this is a task I shall only undertake in the second part of this dissertation. Meanwhile, we have still much to learn—in the first part of the dissertation, we shall review and assess a wide array of analytic theories of conditionals that have been put forward as rivals to the material conditional since approximately the middle of the 20th century. I shall try to acquire some insight into the valid intuitions that gave rise to these theories, while identifying the source of their difficulties or outright failures. My purpose is to glean from this analysis a clear picture of the pre-theoretical properties of natural language indicatives and subjunctives that any reasonable theory should strive to account for.

Subjunctives were obviously much less amenable to the material treatment than indicatives and hence from the very outset invited a radically different approach. They entered the spotlight with Nelson Goodman's and other cotenability theorists' abortive attempt at analysing counterfactuals in terms of lawlike regularities. This approach invited some followers, but never quite attained the popularity of Stalnaker-Lewis ordering semantics, which has been dubbed as 'pretty much everybody's favourite theory of counterfactuals'. However, we shall also take a look at some alternatives. William G. Lycan has attempted to tweak ordering semantics by means of replacing possible worlds by events. Angelika Kratzer's premise semantics based on the notion of lumping has been shown to be mutually intertranslatable with Lewis's system; it nevertheless differs from the latter in its theoretical underpinnings, which are actually closer to cotenability accounts. Kratzer's work has been very well received especially among linguists and linguistically minded philosophers and there has been a number of adaptations of her ideas both in analysis of subjunctives and indicatives. Some of them have sought common ground with the so-called intervention semantics, based on the idea of causal networks. David Sanford's analysis in terms of 'patterns of dependence' bears relation to intervention semantics as well as to Goodman's approach based on cotenability.

As for natural language indicatives, we have already witnessed the resilience of the material interpretation, bolstered by the ingenuity of Paul Grice's theory of implicature. Nevertheless, the persistent difficulties of this treatment led to the development of an alternative 'probability conditional', which exploited an early insight by Frank Ramsey. This has since become something of a mainstream approach to indicatives, championed by such philosophical heavyweights as Ernest Adams, Dorothy Edgington, Jonathan Bennett and others. However, rival theories also abound. Stalnaker intended his similarity-based conditional to work equally well for indicatives as for subjunctives—these hopes sustained an important

blow from Lewis's triviality results. Lycan's theory was originally designed for indicatives. Christopher Gauker's semantics of conditionals makes use of the idea of abstract contexts, conceived as set-theoretical constructs on a universe of literals. Kratzer's restrictor approach, though devised with the eye on improving on Lewis's theory of counterfactuals, was also meant to be applicable to indicatives. Victor Dudman has formulated what has been called the relocation thesis, related to Adams's take on subjunctives and partially subscribed to by Edgington, according to which would-subjunctives differ from will-indicatives only regarding their tense.

And this is only a selection of the most influential theories: there is still more (say, Peter Gärdenfors's inductive logic) that simply had to be left out for the sake of brevity.

All in all, perhaps the most bewildering thing about this panorama are the striking differences between the proposals. How can there be such a rich variety of treatments of what is essentially an everyday linguistic phenomenon? I shall argue that despite appearances, there is much less disagreement about the fundamental issues than meets the eye. Each of these proposals exploits quite a few accurate intuitions about our use of conditionals—the danger hides in overstating their explanatory power. Surprising though it may sound, I submit that we do not have to throw in our lot with any of these analyses, provided we are capable of taking them for what they're worth and nothing more. In second part of the dissertation, I shall try to explain their valid insights by answering crucial questions about the role of conditionals in our intellectual life.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Part I

The theories

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 2

Ramsey Test

2.1. Ramsey and some basic probability notions

One of the first authors to advocate an approach to conditionals radically opposed to explanations based on Frege-Russell's material conditional was F. Ramsey. His simple account of how indicatives are evaluated in practice, which was at loggerheads with the assumption that there was nothing more to it than a mere truth table, has inspired a whole host of scholars of diverse theoretical hew and has been dubbed 'Ramsey Test' in his honour. For Ramsey, disputing about a conditional amounts to disputing about what the degree of belief in the consequent given the antecedent should be:

> If two people are arguing 'If p will q?' and are both in doubt as to p, they are adding p hypothetically to their stock of knowledge and arguing on that basis about q; so that in a sense 'If p, q' and 'If p, $\neg q$' are contradictories. We can say they are fixing their degrees of belief in q given p. If p turns out false, these degrees of belief are rendered *void*. If either party believes $\neg p$ for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses.
>
> (Ramsey, 1929), p.143

The outline of the argument should be clear—if 'if p, q' were a material conditional, ascertaining the antecedent's falsity would suffice to settle the question about its truth. Since it obviously doesn't, the analytical hypothesis must be spurned.

It doesn't seem to me that Ramsey regarded this brief remark as a mature piece of semantic theory. He rather appears to have meant it as

an intuitive description of actual practice lending support to his theory of causal laws—otherwise he would probably not have added it as a mere footnote in an unpublished paper. Note also that it only covers the cases where the antecedent is either regarded as indisputably false or else its truth value is in question. And neither does it as much as hint at how this hypothetical addition of the antecedent to the stock of one's beliefs should look in detail—the idea of a minimal conservative revision that many have read into it is at best implicit. And what about the case where you reject the antecedent without being completely certain about $\neg p$? Should the addition look different in the indicative and in the subjunctive case? These are only a few crucial questions that the original formulation of the Ramsey test was not devised to answer.

It is therefore more than a little bewildering that so many authors should have turned Ramsey's insight (which was not put forward as a proposal about the *meaning* of conditionals) into the cornerstone of a new conditional semantics. We will inspect Stalnaker's treatment of indicative conditionals in Section 4, reserving this one for accounts based on conditional probability. But before proceeding, we have to fix some terminology.

A classical probability space K is defined as a set of state descriptions with a defined probability measure. A state description represents simply a distinct way things may turn out to be. The degree of granularity of a probability space varies with the requirements of the modelled epistemic situation. State descriptions with the maximum degree of granularity will be called 'possible worlds'—they can provide an answer to all imaginable factual queries. Sentences of the language correspond to subsets of the probability space and can be assigned probabilities in accordance with Kolmogorov's probability axioms. They imply, for instance, that $p(\top) = 1$, $p(\bot) = 0$ and for all ϕ, $p(\neg\phi) = 1 - p(\phi)$. If K comprises only a finite number of state descriptions, the probability of any sentence will equal the sum of the probabilities of the state descriptions in the corresponding subset of K. If K is infinite, only its subsets will in general be apt to bear non-zero probability (any state description will in general have zero probability).

Conditional probability of A given B, $p(A/B)$, is defined as $p(A \wedge B)/p(B)$. It represents the likelihood of A should B turn out to be the case. If $p(B) = 0$, then B cannot turn out to be the case under any circumstances and $p(A/B)$ will remain undefined. Even though the definition may suggest otherwise, $p(A/B)$ and $p(A \wedge B)$ are conceptually on a par, and it is often easier to ascertain $p(A/B)$ and derive $p(A \wedge B)$ from it. It was Ramsey's very idea about conditional probability that was to play a central role in Adams's probabilistic approach to conditionals.

2.2. Adams's probability logic

E. Adams elaborated Ramsey's original idea into a full-fledged treatment of indicative conditionals in the course of the 1960s. Its complete presentation was given in (Adams, 1975) and an updated version was presented in (Adams, 1998). Among the many champions of Adams's theory, D. Edgington, J. Bennett and K. A. Appiah stand out.

Adams, as opposed to Jackson, took Ramsey's insight to its ultimate consequences. If it is true that people endorse or reject indicatives based only on the probability they award to the consequent given the antecedent, then why should we ever expect probabilities of such conditionals to be anything else than conditional probabilities? What happens if take the Ramsey test at face value and treat the equation

$$p(A \rightarrow C) = p(C/A) =_{def} \frac{p(A \wedge C)}{p(A)}$$

as a basic fact about indicative conditionals?[1]

Before exploring the consequences of such a decision, some terminology is in order. I shall call the right-hand formula, following Hájek and Bennett, *the ratio formula*. The whole equation will henceforth be referred to, following Bennett, simply as *the Equation*.[2] An alternative, and frankly a little corny label for the Equation you may also find in the literature is *Conditional Construal of Conditional Probability* or CCCP[3]; as well as the more sober *the (Hypo)thesis*.[4] *Stalnaker's (Hypo)thesis*, for reasons that will become obvious shortly, is an interpretation of the Equation that assumes that the probabilities of conditionals thus defined should otherwise behave just as any other unconditional probabilities. *Adams's thesis*, on the other hand, construes the ratio formula merely as a measure of a conditional's *assert(a/i)bility*.

To make matters worse, neither Stalnaker nor Adams subscribe to 'their' respective theses any more—Stalnaker put forward his one in (Stalnaker, 1970), only to discard it upon the publication of Lewis's triviality results, while Adams toyed with 'Adams's thesis' as a rendering of the Equation in his early (Adams, 1965), but later (that is, as of (Adams, 1975)) settled for simple 'probability', construed as a measure of 'reasonableness of belief'[5],

[1] Note that according to this equation, $p(A \rightarrow C)$ remains undefined for $p(A) = 0$.

[2] (Bennett, 2003), p.58

[3] For younger non-Russian speakers, 'CCCP' is the Cyrillic initialism for the USSR. This seems intentional; hence the corniness.

[4] (Hájek and Hall, 1994), p.75-6

[5] (Hájek, 2012), p.147

even though he continued to deny that such probabilities should satisfy Kolmogorov's axioms. Partly to defend his terminology, Adams supplied an independent rationale for accepting the Equation, based on Jeffrey's decision theory. Very roughly, the argument shows that decision makers are best off in the long run if their degree of credence in a conditional equals the conditional probability of its consequent given its antecedent.[6]

And now for the real thing. The immediate upshot of taking the Equation seriously is that natural language indicatives cannot be material conditionals, since the probability of a material conditional can diverge widely from the conditional probability of its consequent given its antecedent. This is because the uncertainty[7] of a horseshoe equals $p(A \wedge \neg C)$ while the uncertainty of an indicative conditional on this account equals $p(\neg C/A)$, that is $p(A \wedge \neg C)/p(A)$. Since $p(A) \leq 1$, the former uncertainty will never exceed the latter and thus $p(A \supset C) \geq p(A \rightarrow C)$. This argument also shows that the only way for the probability of a material conditional to equal the conditional probability of its consequent given its antecedent is for $p(A)$ to equal 1 or $p(A \wedge C)$. Hence $p(A \supset C) = p(A \rightarrow C)$ only in the extreme cases when the antecedent is perfectly certain itself or implies the consequent logically.

As the truth conditions of the material conditional can't give us what we want, it would seem that the obvious next thing to do is to look for alternative ones, ones that would guarantee the probability of the indicative conditional to match the conditional probability of the consequent given the antecedent. Yet we will see in the next section that D. Lewis proved, with the help of only a handful of apparently harmless assumptions, that in general no such truth conditions can be found. The power of these results has led Adams, as well as other authors who have embraced the ratio formula, to conclude that the probability of indicative conditionals doesn't amount to the probability of their truth and that indicatives thus cannot be called true or false (or, for that matter, 'propositions') in the standard sense of the word.[8]

Be it as it may, if for one reason or another we are not in a position to formulate truth conditions for indicative conditionals, we run into a funda-

[6] See Chapter III of (Adams, 1975) and Chapter 9 of (Adams, 1998).

[7] Uncertainty of a formula is defined as the rate of its improbability, or one minus its probability.

[8] (Adams, 1998), p.199. Now, although I am sympathetic to the probabilistic treatment of indicatives, and certainly wouldn't want to downplay the importance of the triviality results, I believe this conclusion should be put into perspective. As well as in the next section, I will take this subject up in Section 6.3.1 when I address the issue of subjectivity of indicatives.

mental problem concerning their logic. If we wish to formalise their inferential behaviour within a deductive system, the standard truth-conditional notion of validity will no longer be up to the job.

Remember that a valid argument is standardly defined as one whose premises cannot be true without the conclusion winding up true as well. Model theory supplies a mathematically precise way of cashing in this informal definition. Our problem here comes down to the realisation that indicative conditionals are either not properly conceived of as true or false in a model, or else their models should be of an altogether different kind than those of simpler (or, on this view, 'factual') propositions. Either way, the usual notion of validity will be inapplicable to arguments combining indicatives with simple propositions. This is why Adams comes up with his probabilistic criterion of validity instead:

Probabilistic Entailment A set of formulas Θ probabilistically entails ψ iff the uncertainty of ψ cannot exceed the sum of uncertainties of all formulas in Θ.

Here, the modal 'cannot' should be interpreted as making tacit reference to probability assignments satisfying Kolmogorov axioms and proper to all formulas in Θ together with ψ.

This definition generalises classical truth-conditional validity for arguments involving probabilistic conditionals. That is, if there are only classical formulae among Θ and ψ, then Θ implies ψ classically if and only if it also does so probabilistically. However, as soon as conditionals make their entrance, the former notion of validity is no longer available and the probabilistic one must be applied instead.[9]

The combination of the ratio formula and the probabilistic notion of entailment manages to throw light on the most egregious paradoxes of the material conditional. For instance (CAC),

(38) $C \therefore A \rightarrow C$

[9]There is some superficial vacillation about this definition in Adams (see Bennett's remarks in (Bennett, 2003), p.129). The 'probabilistic soundness criterion', as announced at the very beginning of (Adams, 1975) and evoked a few times throughout Chapter I, requiring that the premises of a valid argument cannot be probable while its conclusion is improbable, is out of step with the definition given here, which in turn is equivalent to the one put forward in Chapter II (p.57) and explicitly adopted by Adams in (Adams, 1998), p.131. The bottom line is whether we would like to label arguments such as the one from the famous Lottery Paradox as probabilistically invalid. Since the latter option only brings limited benefits at the huge cost of dissociating the two soundness criteria as the number of premises grows, the definition championed here has generally been found preferable.

is probabilistically invalid, since no matter how probable (shorn of complete certainty) C comes out on a given probability distribution, $p(C/A)$ may still be very low, provided most of A lies outside C. This requires that A should itself be very improbable, which agrees with intuition—typical examples that bring out the paradoxicality of (38) include the likes of

(39) Tomorrow is Friday. Therefore, even if I've got mixed up with the dates, tomorrow is Friday.

The same goes for the complementary pattern (NAAC)

(40) $\neg A \therefore A \rightarrow C$,

since unless the possibility A is completely excluded, the probability of $A \rightarrow C$ can be arbitrarily low despite $\neg A$'s likelihood. Now it is the very premise that requires that the antecedent should be very improbable for the counterexample to go through:

(41) The stores do not open tomorrow. Therefore, if the stores open tomorrow, they won't have any customers.

The invalidity of the above inference patterns is sometimes generically chalked up to the alleged 'non-monotonicity' of probabilistic reasoning. However, we have to be careful, for such formulations can easily lead us to error. Probabilistic validity, as defined, *is* monotonic with respect to the logical strength of the premises. Where non-monotonicity enters the picture is when *new information* is combined with the accepted premises in order for us to draw new conclusions. In view of the fact that such an addition can drastically subtract from the probabilities of the premises, the 'static' notion of validity can only be applied to their *posterior* uncertainties, that is, to the uncertainties of the premises conditional on the new information. This is how Adams accounts for the fact that sometimes highly probable premises coupled with new information entail very unlikely conclusions.

The basic idea is that conditional antecedents work analogously to such new information.[10] They can introduce hypotheses capable of trumping the accepted assumptions. This doesn't make validity non-monotonic, but rather implies that conditional antecedents do not work as 'additional premises' in sustaining the consequent. The correct diagnosis, therefore, should be that the *deduction theorem*, also called *conditional proof*, fails for Adams conditionals: $\Theta, A \therefore C$ doesn't guarantee $\Theta \therefore A \rightarrow C$.

Besides the inferences valid for the material conditional but generally recognised as paradoxical, however, Adams's logic also renders invalid

[10](Adams, 1998), p.142

some patterns that have long been considered to make up the backbone of conditional logic. They are Antecedent Strengthening, Hypothetical Syllogism, Contraposition and Disjunctive Syllogism. While there certainly are examples that cast a shadow of doubt on these inferential patterns, renouncing them faces us with the new theoretical challenge of explaining the origin of their unmistakable halo of validity. I shall take this topic up in Section 6.3.3, where the issue of alleged polysemy of our indicatives will be addressed.

2.3. Truth conditions and triviality

2.3.1. Lewis's proofs and their legacy

No exposition of probability as applied to conditionals can be complete without touching upon the triviality results which we briefly mentioned in the preceding section. The original proof was given in D. Lewis's 'Probabilities of Conditionals and Conditional Probabilities'.[11] The bottom line of Lewis's argument was that attempts to formulate standard truth conditions of indicatives in a way that would make their absolute probability conform to the ratio formula, were doomed to failure. Since Lewis's first paper, a whole industry in triviality proofs has taken off, with the aim of proving the same or analogous results with simpler or slightly modified assumptions.

Perhaps the most notorious proposal equating probabilities of conditionals with conditional probabilities, and certainly the one that spurred Lewis to come up with the triviality results, was put forward in (Stalnaker, 1970). In this article, Stalnaker suggested that compliance with the equation should be regarded as a minimum requirement on a reasonable account of truth conditions of conditionals (at that point he hadn't yet delved into the distinction between indicatives and subjunctives). He believed his own treatment satisfied this criterion, while Lewis's couldn't.[12] Lewis showed that such requirement had some extremely unpalatable consequences.

The idea behind Lewis's reasoning is quite simple, but not less brilliant for that matter. If Stalnaker is right and $A \rightarrow C$ is a regular proposition, that is, it divides the probability space (conceived as a set of state descrip-

[11] (Lewis, 1976)

[12] The reason was that the material interpretation of indicatives was obviously at odds with the equation and Lewis's logic for counterfactuals didn't validate Conditional Excluded Middle (CEM), implied by the Equation. See (Hájek and Hall, 1994), p.77.

tions or possible worlds) into two regions according to its truth value in each of them, then its probability must comport with standard probability theorems. In particular, it should satisfy expansion by cases:

(EbC) $p(A \to C) = p(A \to C/B) \times p(B) + p(A \to C/\neg B) \times p(\neg B)$

for any proposition B. Moreover, if we assume, as seems reasonable, that the meaning of a conditional is preserved between p and the probability function p_B obtained from p by conditionalising on B, then the Equation also guarantees that the following reduction principle should hold:

(RP) $p(A \to C/B) = p(C/A \wedge B)$

In other words, the conditional probability of a universal probability conditional given another proposition should equal the conditional probability of the consequent given the antecedent and this other proposition. Now all ingredients are in place to derive the paradox—it suffices to take any conditional $A \to C$ that is contingent (i.e. with neither $p(A \wedge C)$ nor $p(A \wedge \neg C)$ equal to 0) and expand it with respect to its very consequent:

$$\begin{aligned} p(A \to C) &= p(A \to C/C) \times p(C) + p(A \to C/\neg C) \times p(\neg C) \\ &= p(C/A \wedge C) \times p(C) + p(C/A \wedge \neg C) \times p(\neg C) \\ &= 1 \times p(C) + 0 \times p(\neg C) \\ &= p(C) \end{aligned}$$

This means that irrespective of our choice of A and C (under the specified conditions), they turn out to be probabilistically independent. Lewis proved that this can only work for *trivial* languages, which lack the resources to express more than two incompatible sentences with non-zero probability.

The above informal presentation of the argument, corresponding to Lewis's first triviality result, takes for granted that our conditional should satisfy the Equation across all possible probability functions. As this requirement may justifiably seem excessive (not all probability functions represent viable belief systems), Lewis later generalised the result for probability conditionals for any class of probability functions closed under conditionalisation on the cells of finite partitions of the probabilistic space as well as for any class of probability functions closed under Jeffrey

conditionalisation.[13]

Stalnaker's project was proved inviable by the very first triviality result and he didn't hesitate a second to jettison it upon acquainting himself with Lewis's work. Philosophically relevant attempts to resuscitate a truth-conditional account of indicatives that would comport with the Equation have been extremely scarce since. However, van Fraassen's work constitutes an interesting exception—in a much-discussed article[14], he called into question Lewis's assumption that conditionals should preserve their meaning across probability assignments, which he chalked up to Lewis's metaphysical realism.[15] According to van Fraassen, distances between possible worlds should not be construed as fixed features of an independently existent reality, but only as representations of belief revision policies. It is therefore not unreasonable to admit that a change of a probability assignment can impinge on the extension of a Stalnaker conditional. This move undercuts Lewis's proof, as the uniformity assumption is required for (RP).

Against this, Stalnaker (at that point his suspicion of the 'Stalnaker's thesis' measured up to his earlier confidence in it) proved that the Equation entailed triviality with the help of his C2 logic and free conditional embedding even in the absence of Lewis's uniformity assumption.[16] Hájek in turn proved, in what he later called the 'wallflower argument'[17], that in any finite model (of cardinality greater than 2) conditional probabilities strictly outnumber unconditional ones, hence there must be at least some conditional sentences whose probability doesn't equal the ratio formula.[18]

Many other mathematically minded philosophers also rose up to the challenge of exploring how far triviality could be pushed. You could for instance further relax the requirement on the conditional to preserve its meaning across probability assignments or argue that neither ordinary nor Jeffrey conditionalisation are reasonable models of belief revision. However, it has been proved by N. Hall that even such drastic manoeuvres cannot ward off triviality. Unless two probability functions are orthogonal (that is, they respectively assign perfect certainty and uncertainty to some

[13] See (Lewis, 1986b). Despite the relaxed assumptions, the guiding idea behind the proof of triviality results two to four remains basically the same as in the first one.

[14] (van Fraassen, 1976)

[15] See Section 4.3.1.

[16] See (Stalnaker, 1976). As van Fraassen championed a different logic than Stalnaker, his own proposal was not directly affected by Stalnaker's argument. However, Stalnaker had originally used the Equation only as an argument in favour of C2 and he was more loath to give up the latter than the former.

[17] (Hájek, 2012), p.156

[18] (Hájek, 1989)

sentence), Stalnaker's thesis must fail at least for one of them on pain of triviality.[19] If you think that no reasonable model of belief revision can turn a probability assignment into an orthogonal one, you had better give up Stalnaker's thesis.

All these problems have led many scholars, including Stalnaker, to frown upon the Equation as their ultimate source. Remember, however, that when we defined Stalnaker's thesis, we characterised it as a particular interpretation of the Equation, which combined it with a propositional construal of indicatives. The latter assumption guarantees that probability assignments to compound formulae with embedded conditionals will exist and satisfy Kolmogorov's axioms, which is a usual requirement for a triviality proof to go through. Even Hájek's proof, frugal as it is with premises, only keeps its edge if you want to assimilate conditionals to non-conditional propositions.

However, there is another way of looking at things. P. Egré and M. Cozic have come up with a particularly helpful interpretation of triviality results—perhaps they should rather be viewed as showing the undefinability of conditional probability operators in terms of unary probability operators together with binary propositional connectives.[20] From this vantage point, indicative conditionals, as bearers of conditional probability, broaden the expressive power of our vocabulary beyond what could be said with only simple propositions.[21] This opens a way out for the champions of the Equation—retain the latter and blame all the trouble on the auxiliary part of Stalnaker's thesis, which equated the content of conditionals to that of simple propositions. D. Edgington has argued convincingly that this is the right way to go.

2.3.2. Edgington against truth conditions

In her (Edgington, 1986), Edgington put forward an argument, whose main inspiration bears striking similarity to Lewis's proofs, in favour of the thesis that indicative conditionals lack (non-conditional or otherwise

[19] (Hájek and Hall, 1994), p.97.

[20] See (Egré and Cozic, 2011); a similar point is also made in (Edgington, 2007), p.206. Of course, Cozic's and Egré's actual proposal follows Lewis-Kratzer's restrictor approach to indicatives, striving to bestow truth conditions on them while at the same time endorsing the Equation (see Section 5.2). This, however, doesn't prevent one from using their clever idea for different theoretical purposes.

[21] This is reminiscent of Brandom's dictum to the effect that with the introduction of conditionals, one can explicitly say what she otherwise only could display in her inferential behaviour. See for instance (Brandom, 2001), p.60.

non-trivial) truth conditions. In this context, this amounts to the contention that indicative conditionals do not divide the probabilistic space into two disjunct regions according to their truth value in each state description. Their task is to say something about the probability space as a whole, not about any particular possible state or world.

The material conditional being the only sensible truth-functional candidate for the meaning of indicatives, Edgington sets off by dismissing the case for it, using arguments that will have become familiar to the reader by now. She then goes on to argue that any attempt at a truth-conditional but non-truth-functional account is bound to slip back to the material interpretation. The bottom line is that if indicatives have truth conditions, they must be those of material conditionals. As these are obviously at odds with the meaning of indicatives, we are forced to conclude that the latter lack truth conditions altogether.

What does it mean for an account to be truth-conditional without being truth-functional? Such an account attributes to the conditional a meaning akin to that of 'because': for 'ψ, because ϕ' to be true, it is necessary but not sufficient that both ψ and ϕ be true. Applied to a conditional, there must on such an account exist at least one line of the truth table for A and C in which $A \rightarrow C$ can consistently have two different truth values. If we cannot find such a line, we are back in the mire of the original truth-functional account. The following table brings out the situation:

	A	C	$A \supset C$	C, because A	$A \rightarrow C$
1.	T	T	T	T/F	?
2.	T	F	F	F	?
3.	F	T	T	F	?
4.	F	F	T	F	?

Edgington enjoins us to imagine we are in a state where the truth values of the line in question obtain and ask ourselves whether we know enough to decide about the indicative. This is straightforward for the second line (with A true and C false, $A \rightarrow C$ must be false) and almost equally so for the first line (with both A and C true, $A \rightarrow C$ arguably cannot fail to be true as well—remember that on the level of particular states/possible worlds there is no room for uncertainty. So 'even if A, C' seems like a safe bet.)

The situation is more complicated with lines 3 and 4. We have no clear intuitions about $A \rightarrow C$ when A is false, as predicted by the Equation. Edgington argues that given the truth of the consequent on line 3, we should be more inclined to judge the whole conditional true, but this is

no knock-down argument. She proceeds rather by making use of the previously established result, asking us instead to imagine we are in a situation where we are uncertain between line 3 (or 4) and 1. This seems to be enough for perfect certainty about the conditional, which, together with the assumption of truth-conditionality, implies we have to regard the conditional as true also on lines 3 and 4. But then we wind up with the truth table of the material conditional and our conditional turns out to be truth-functional after all.

That we are asked to combine lines 3 and 4 with line 1 only has to do with the structure of Edgington's broader argument, as outlined above, and lacks deeper significance. We could of course combine them with line 2 in order to obtain exactly the opposite result. This poses no difficulty to Edgington's thesis—it is only because indicatives are not properly evaluated at the level of state descriptions (as any respectable 'proposition' must) that we can end up with conflicting intuitions about their truth values at some of them. If indicatives had truth conditions, the argument goes, our judgement would not hinge upon which other state descriptions are compatible with our knowledge.

This argument is presented in an even more succinct and poignant form in (Edgington, 2007):

> Two prima facie desirable properties of indicative conditional judgements:
>
> (I) Minimal certainty that $A \vee B$ (ruling out just $\neg A \wedge \neg B$) is enough for certainty that if $\neg A$,B; [...]
>
> (II) It is not necessarily irrational to disbelieve A yet disbelieve that if A,B.
>
> The truth functional account satisfies (i) but not (ii). Stronger truth conditions may satisfy (ii), if they allow that the conditional can be false when A is false. But they cannot satisfy (i): for any stronger truth condition, ruling out just $A \wedge \neg B$ leaves open the possibility that 'If A,B' is not true.
>
> (Edgington, 2007), pp.169-170

The upshot is that if you conceive of indicatives on a par with non-conditional propositions, Edgington's (i) will require them to be at most as logically strong as material conditionals, whereas (ii) will pull in the exact opposite direction—if the antecedent's lack of probability cannot guarantee the probability of the whole conditional, then neither can the mere falsity of $A \wedge \neg B$.

However, you can keep both (i) and (ii) provided you maintain your allegiance to the Equation, but reject the rest of Stalnaker's thesis. This is the way out championed both by Adams, Edgington and Bennett. An indicative, on such an account, does not introduce yet another piece of information, independent of the truth of its antecedent and consequent. Rather than that, it is a device tasked with the management of imperfect information. Assignment of fixed truth values to the operands of an indicative conditional, which occurs at each possible world or state description, takes away the uncertainty that is vital to its job description, thus rendering it useless. That's why conditional probability is all there is to say about the acceptability indicatives and that's also why their meaningful embeddability within compound sentences turns out to be severely restricted.

2.4. Conclusion

We have witnessed the many advantages of a probabilistic approach to indicative conditionals. However, they should not make us jump to conclusions—it is all to easy to get hold of the wrong end of the stick and convince ourselves that in view of Adams's and Edgington's arguments, indicatives are correctly conceived as mere devices for expressing subjective conditional probabilities and cannot therefore be properly called true or false.[22] In Chapter 6 we will see why such a construal of the results we have reviewed here is riddled with difficulties. We shall also examine a simple alternative, which is compatible with the insights about conditional probability, but doesn't give rise to the problems besetting the subjectivist construal. However, before that, we need to devote some attention to the treatment *subjunctive* conditionals have received in the literature.

[22]This is the view championed, among others, by A. Gibbard in (Gibbard, 1981) and J. Bennett in (Bennett, 2003).

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 3

Cotenability

The first analytic theories of subjunctives appeared in the 1940s and were pioneered by authors such as Roderick Chisholm, Nelson Goodman and Nicholas Rescher. I will dedicate more space to them than is customary in contemporary essays on the topic, chiefly because I'll try to show that later analyses, often unwittingly, had to face very similar challenges. Sound understanding of the technical difficulties encountered by early writers on subjunctives will give us a handle on the solutions put forward by their successors. Another reason is that the early scholars tried to connect the meaning of subjunctives with their actual use in rational practice—a commendable attitude that, inexplicably, has been all but lost since the advent of modern similarity-based possible-worlds semantics.

A recurring theme among the first writers on subjunctives is the defence of the latter as meaningful and valuable pieces of discourse, against suspicions from some corners of analytic philosophy. The qualms of an empiricist can easily be understood: counterfactuals seem to make statements about unrealised possibilities and as such lack direct empirical testability. What they convey is not expressible in classical logical formalism, as the material conditional is blatantly inadequate for this task (coming out true whenever its antecedent turns out to be false). And apart from not being truth-functional, as Quine famously remarked, many of them seemed to resist motivated evaluation altogether, which cast a shadow of doubt on the very possibility of a systematic theory of natural language counterfactuals[1]. The most pressing task for early authors on counterfactuals was therefore to show that they featured widely in respectable discourse and were thus not a mere product of imagination-driven daydreaming.

We can see examples of this both in Chisholm and Rescher. The former

[1](Quine, 1966),p.14-15

claims, for example, that

> [...] adequate understanding of science and history requires the ability to consider the consequences of hypotheses known to be contrary-to-fact.
>
> (Chisholm, 1946), p.290

According to Chisholm, counterfactual hypotheticals are essential in precautionary and deliberative reasoning and they supply the grounding for successful applications of dispositional predicates. They are also widely used in philosophical analysis, even by authors that might be regarded as the most reluctant to admit them, such as logical positivists. Rescher makes similar points and, in particular, he underscores the crucial role of the possibility of entertaining 'belief-contravening suppositions' in reasoning by *reductio ad absurdum*.[2]

Once counterfactuals acquired analytic philosophy's seal of approval, such forays into their natural habitat have been mostly regarded as superfluous and excised from papers and books on the topic. I for one think that's a pity. After a few decades of focusing on a fixed as well as circumscribed set of examples, we run a great risk of losing sight of *when* and *why* we actually tend to employ counterfactual reasoning, which is as likely as anything to provide a clue to the meaning of subjunctives.

Many scholars follow D. Lewis[3] in labelling these early approaches as 'metalinguistic theories'. They agree on tracing back a counterfactual's truth (or, in case of non-truth-conditional accounts, assertibility) in one way or another to the validity of a certain implicit argument, which supplies the connection between the antecedent and the consequent. On this account, a counterfactual makes a metalinguistic statement to the effect that this argument is valid. We can represent the latter schematically as

$$A, \psi_1, \psi_2, ..., \psi_n \therefore C^4,$$

where $\psi_1, \psi_2, ..., \psi_n$ stand for some true factual statements that function as collateral premises.[5]

---

[2](Rescher, 1961), p.178

[3](Lewis, 1973a), p.65-77

[4]Since for the authors we are considering in this section 'validity' means 'logical validity', '$\therefore$' should be read as 'entails'.

[5]Most early theorists did not draw a clear distinction between propositions and their linguistic expression, so they did not allow for the possibility of an utterer tacitly referring to matters of fact inexpressible in his language. This was remedied in Kratzer's premise semantics.

The exact nature of the connection between the counterfactual and the corresponding argument is construed slightly differently by each author. For most of them (Chisholm, Goodman, Rescher), the counterfactual only asserts that *there is* such an argument, yielding what we shall call *the basic schema* of metalinguistic theories:

Basic Schema MT $A > C$ is T iff there are propositions $\psi_1, \psi_2, ..., \psi_n$ satisfying the condition Φ, such that '$A, \psi_1, \psi_2, ..., \psi_n \therefore C$' is a logically valid argument.

However, for Mackie a counterfactual is itself an abbreviated, 'telescoped' version of such an argument and does not express a higher order statement.[6] The upshot is that for him, counterfactuals are neither true nor false, but rather valid or invalid (or acceptable/inacceptable) depending, of course, on the validity of the unabridged argument. Since I believe that the issue of whether a typical counterfactual makes a metalinguistic claim or not is secondary to the problem of restricting the set of collateral premises, I have opted for Goodman's term 'cotenability' in order to characterise the early theories of subjunctives.[7]

This approach to counterfactual conditionals roughly follows the pattern of C.I. Lewis's treatment of strict implication, which he also conceived as making what we now would call a meta-linguistic statement and thus differing radically from Russell's and Whitehead's material conditional, a mere truth-functional operator. The strict implication $A \dashv C$ was introduced to codify the relation of entailment between A and C, without any additional premises:

$A \dashv C$ is T iff '$A \therefore C$' is a logically valid argument.[8]

So when $n = 0$ in the Basic Schema, the connection between A and C will be purely logical and the counterfactual will degenerate to a strict implication. Actually, the fact that some implications with false antecedents (which in natural language would arguably be expressed by means of

[6] See (Mackie, 1962), p.68 and (Mackie, 1973), p.69.

[7] Moreover, there is certainly nothing metalinguistic in the current revival of this kind of analysis by A. Kratzer, as we shall see in Section 5.2.

[8] See (Kneale and Kneale, 1962), p.554. In this connection, it is worth pointing out that both Goodman and Chisholm were C. I. Lewis's graduate students at Harvard. We shouldn't forget, however, that Lewis had a proof-theoretic notion of entailment in mind (see, for example, (Korte et al., 2009), p.532). Lewis's work follows the syntactic paradigm of *Principia Mathematica* and as such precedes Tarski's distinction between syntactic and semantic consequence.

counterfactuals) are acceptable and others not was one of the arguments C. I. Lewis advanced to show the unfitness of material implication to codify the information these implications convey.[9]

This explains why $A \dashv C$ entails $A \boxdot\!\!\rightarrow C$. However, more often that not, logic alone will not suffice to bridge the gap between the antecedent and the consequent of a counterfactual—some additional mortar must be brought in so as to make them stick together. Consider Goodman's example

(42) If this match had been struck, it would have lighted.

If (42) sounds acceptable, it seems that this is only thanks to certain facts about the match (e.g. the chemical composition of its head, dryness, etc.) and its environment (e.g. the presence of oxygen, etc.), as well as a general principle that connects scratching such a match under the given conditions to its lighting. As these informal considerations suggest, we can draw a rough-and-ready distinction between two kinds of supporting statements. Most scholars have distinguished between

1. *Relevant conditions*: actual matters of fact that have to be added to the antecedent in order to infer the consequent.

2. *Laws*: general connecting principles that shore up the inference from the relevant conditions and the antecedent to the consequent.

As a result, a satisfactory analysis of counterfactuals should explain a) how relevant factual conditions are selected, and b) how to distinguish nomological generalisations, capable of supporting counterfactuals, from merely accidental ones. Each of the early theories had to address these two problems—we will take on them one at a time.

3.1. Relevant conditions

Given a subjunctive, the problem of relevant conditions concerns the restriction of the set of actual facts eligible as additional premises in the argument that backs it.[10] It is not difficult to see why a metalinguistic theory must impose such a restriction. After all, in a prototypical case the

[9] See (Lewis, 1912), p.529.

[10] In (Bennett, 2003), the term 'relevant conditions' is rejected as a misnomer, because 'irrelevant conjuncts [...] cannot lead to any conditional's being accorded a truth value that it does not deserve.'(p.307). The crucial point is rather to keep *illegitimate* facts out of the set of additional premises and that's why Bennett prefers to call this set 'Support'. This reflects a fundamental difference in the spirit of metalinguistic approaches on the

set of all facts will include the negation of the antecedent, which certainly should not be let in, on pain of making all counterfactuals trivially true (since an argument with contradictory premises is valid regardless of its conclusion). While everybody agrees on this much, how to carve the set of relevant conditions further is matter of some dispute.

3.1.1. The liberal wing: Chisholm, Rescher and Mackie

Perhaps the most liberal stance on the issue of relevant conditions was taken by some of the earliest writers on counterfactuals . Consider the following pair of examples:

(43) If Apollo were a man, he would be mortal.

(44) If Apollo were a man, some man would be immortal.

Chisholm observes that there is no *prima facie* reason to prefer one over the other, even though they cannot be asserted in the same breath. We could take on board as collateral premise either what we know about Apollo's immortality (a relevant condition) or what we know about human mortality (arguably a lawlike generalisation). According to Chisholm, it is the context of enquiry that dictates the proper choice.[11]

But are there any bounds to what we can take in as collateral premise? This is a crucial point, and Chisholm opts for the simple answer: there are none. He suggests that the speaker has virtually complete freedom over the choice of statements that back her counterfactual.

At first, Chisholm conjectured that this was usually due to antecedent ambiguity, which had to be pragmatically resolved through a process of recovery of the speaker's meaning. Later he slightly changed the picture, bestowing different roles on the antecedent and the supporting premises. He called the former *supposition* and the latter *presuppositions*. Presuppositions no longer simply expand the supposition—their role is different, as they are *assumed* rather than merely *entertained* (as the supposition is). Sim-

one hand and world-based approaches on the other: whereas an ideal analysis of the first kind would positively specify which actual facts support a counterfactual, an analysis of the second type will content itself with smuggling them in along with a plethora of other, irrelevant facts. Bennett is basically right, and Goodman's strategy of restricting the set of additional premises instead of building it up from the bottom bears him out on this. However, since there is little risk of confusion, in deference to the history of the subject, I stick with Goodman's terminology.

[11](Chisholm, 1946), pp.304-305. The mythical nature of the example is not meant to have any bearing on the analysis.

ply making the presuppositions explicit in the antecedent would therefore distort the import of the original counterfactual.[12]

The main point still stands, though—the recovery of presuppositions is still crucial for resolving the vagueness of a counterfactual. Admittedly, if the listener doesn't work out what the speaker's choice of presuppositions is, the communication will break down. However, and crucially, this will be due to the pragmatic inefficiency of the speaker and not to any genuine disagreement. There is simply no right selection of supporting premises.

The very liberalism of Chisholm's picture makes it urgent to explain how the tacit understanding between the speaker and her audience, which determines the background against which a counterfactual is evaluated, comes about in a particular context of inquiry. But such explanation will be in order even on a more restrictive approach, as long as it admits certain margin for variation. It is to Chisholm's credit that he drew our attention to this phenomenon, even though he didn't even begin to explain it and in all likelihood exaggerated its magnitude.

The leeway Chisholm's theory accords to speaker's communicative intentions certainly seems excessive. Some combinations of collateral premises give rise to inconsistent sets—something that Chisholm explicitly allows. So you can properly assert, while pointing to a wooden stick,

(45) If that were gold, some things would be malleable and not malleable.

This can be achieved simply by taking as supporting truths both the malleability of gold and the non-malleability of wood. Bennett argues that counterfactuals with possible antecedents and impossible consequents are never true and thus always unacceptable.[13] Hence, according to Bennett, at the very least a consistency restriction should be imposed on the set of collateral premises.[14]

Should we impose more restrictions on the set of collateral premises? Rescher largely shared Chisholm's intuitions about the contextual pliability of counterfactuals and attempted to provide an explanation. If our beliefs

[12](Chisholm, 1955), pp.102-103

[13](Bennett, 2003),p.306

[14]However, Chisholm could dodge this objection by admitting that impossible consequents are indeed unacceptable, but on pragmatic rather than semantic grounds—they signal that the premises of the implicit argument are inconsistent, and as we are hardly ever interested in assessing such arguments, the speaker's behaviour would flout the Gricean maxim of relevance. It is not that a speaker could not assert such a counterfactual without saying something false, it is rather that in asserting it he could not make a germane contribution to the conversational exchange.

are $S_1, S_2, ..., S_n$, then what he calls a 'belief contravening supposition' $\neg S_1$, will more often than not be inconsistent not only with S_1, but also with the set of beliefs $\{S_2, ..., S_n\}$ as a whole. Rescher, as opposed to Chisholm, insisted that our counterfactual scenarios must be logically consistent and empirically congruous, so some of the latter beliefs will have to be shed along with S_1. And since there usually will be a number of ways to curb this set so as to make it consistent with $\neg S_1$, logic alone is not in a position to determine which of a competing pair of counterfactuals will be acceptable.[15]

So is there any other criterion that could help us break the tie? There is indeed, according to Rescher, though only for *nomological* counterfactuals—that is, counterfactuals based entirely on a covering law. Since covering laws play a privileged role in our epistemic lives, we are reluctant to give them up and prefer to forgo some factual beliefs instead when faced with a choice. Based on these considerations, we can adjudicate between (43) and (44) after all—Rescher predicts that as (44) contravenes a lawlike generalisation, (43) will be preferred by most speakers. Nevertheless, in the case of *hypothetical* counterfactuals, which do not exemplify covering laws, there is no such criterion and it is the so-called dialectical background that is called for to settle the dispute (a move highly reminiscent of Chisholm's recourse to speaker's meaning). This is exactly the case with the pair which made Quine question the whole enterprise of searching for a theoretical account of natural language counterfactuals:[16]

(46) If Bizet and Verdi had been compatriots, Verdi would have been French.

(47) If Bizet and Verdi had been compatriots, Bizet would have been Italian.

So while adherence to laws is non-negotiable for Rescher, he roughly followed Chisholm's liberal line as far as relevant conditions are concerned.[17]

Mackie, for his part, somehow thought that his non-truth-conditional account of counterfactuals relieved him from the duty of specifying the

[15](Rescher, 1961),p.181-2

[16](Quine, 1966),p.14-15

[17]Almost half a century later, in his (2007), Rescher further developed what remains essentially the same theory. The inconsistency that arises when a counterfactual supposition is added to a set of beliefs is handled partly by appeal to 'fundamentality prioritization', which enjoins us to award precedence to propositions in accordance with the role they play in the system of our beliefs (manifested in the *Meaning-Existence-Law-Fact* hierarchy), and partly by appeal to the 'questioner's prerogative', which breaks ties between propositions of otherwise equal standing. See (Rescher, 2007), pp.89-137.

restrictions on tacit premises very thoroughly.[18] He accepted Rescher's point about laws, but explained it pointing to the inductive grounds on which they are believed. However, as regards relevant conditions, he went with the liberal flow of the time. He argued that cases of ties were just that—ties. Since counterfactuals are never really true or false, there is no *prima facie* hurdle to viewing both of a pair of competing counterfactuals as unfit for use.

Does the liberal version of the metalinguistic account fare better with laws given pride of place? It is highly doubtful. Chisholm, Rescher and Mackie all seem to take for granted that in a conversational exchange, the interlocutor who asserts a counterfactual, or the one who puts forth a counterfactual question, should be able to tighten up the slack in her description of the counterfactual scenario if requested to do so. But that's a tall order. More often than not we do not have the slightest idea of the facts that underpin our counterfactuals, even if we are quite positive about there being such facts. Bennett argues that we can affirm

(48) If I had pressed the button, the red light would have gone on.

in the absence of any knowledge about the wiring that connects the button with the bulb. The meaning of a counterfactual guarantees only that pertinent additional premises exist, not that the speaker is able to produce them on request.[19]

This may not be the best possible example, as we can imagine someone asserting (48) based on a direct lawlike connection between pressing the button and the red light (most of the 'laws' we apply in our daily reasoning are reducible to more fundamental facts and principles). Moreover, Chisholm *does* state explicitly that in asserting a counterfactual, we are only committed to the existence of some adequate collateral premises and not to knowing what they exactly are.[20] We might therefore suspect that what Chisholm really had in mind was that the speaker had to be prepared to strengthen the antecedent (or make his presuppositions explicit, as the case may be) only in case and to the extent that the basic schema produced conflicting results.[21]

---

[18](Mackie, 1962), p.74 and especially (Mackie, 1973), p.69. I consider Mackie's laid-back approach unjustified. You still have a theoretical duty to explain what makes a premiss admissible in the collapsible argument hinted at by a counterfactual, regardless of whether the latter makes an ordinary or a higher-order statement.

[19](Bennett, 2003), p.304

[20](Chisholm, 1946), p.299

[21]However, this interpretation seems to be contradicted by the examples in (Chisholm, 1955), in which the whole underlying argument is invariably made explicit when resolving

This way of shrugging off the problem won't do, however. First, it still places an unrealistic burden on the speaker—few of us could fill in even an incomplete argument form for any counterfactual that we have confidently advanced. This suggests that an analysis along the lines of the basic schema is not something that underlies our usage of counterfactuals, but rather an attempt at a rational reconstruction of this usage. Such reconstructions may play a crucial role in spelling out the truth conditions of counterfactuals, but accounts that presuppose conscious access to them must shoulder a huge burden of proof.

Second, and crucially, if Chisholm's theory were true, counterfactual reasoning would simply boil down to questions about logical validity. This is what Chisholm has to say about counterfactual questions:

> Similarly, if we ask, "What would American policy in Asia be if Stevenson were President", our question is incompletely formulated; a more explicit formulation would be, "Supposing that Stevenson were President, and presupposing so-and-so, but not so-and-so, what would be the consequences with respect to American policy in Asia?"
>
> (Chisholm, 1955), p.104

The implausibility of this approach is reflected by the question-begging flavour of such questions. Chisholm seems to suggest that by raising a counterfactual question, the speaker only invites the interlocutor to do some 'data-crunching' for her and churn out the logical consequences of the premises she has in mind. However, a moment's reflection clearly will show that this picture is unsustainable. Reasoning under counterfactual assumptions is not arbitrary and there are constraints on how the counterfactual scenario is built up. Counterfactual questions, far from being trivial, call on the interlocutor to use *his own resources* to envisage such a scenario.

Third, relegating the problem of collateral premises to the realm of pragmatics doesn't let us off the hook. There are really two issues that need to be addressed—on the one hand, the contextual sensitivity that some counterfactuals such as (43) and (44), or (46) and (47) show, and on the other, the specification of collateral premises suitable for filling in the basic schema. Authors such as Chisholm, Rescher and Mackie seemed to believe that an appeal to speaker's meaning or contextual salience would not only solve the first problem, but also the second one. However, this still leaves us in the dark about the mechanism of selection of the relevant

counterfactual ambiguity. It seems that Chisholm progressively set more and more store by pragmatic, rather than semantic restriction of the set of available premises.

facts by the speaker, as well as about the audience's miraculous capacity of working out the full purport of the speaker's contribution. It behoves us to provide an explanation of these phenomena, regardless of whether they are semantic or pragmatic in nature.

3.1.2. Goodman's take on cotenability

One can only wonder what the fate of Goodman's classic (1947) article would be under the present publishing culture. There are certainly not many authors nowadays that dare to admit to a double failure and still get published—on the contrary, any limp in an analysis tends to be carefully concealed rather than openly acknowledged. The fact that despite the apparent handicap, this paper managed to set the agenda for research into counterfactuals for the decades to come, should give us pause. Maybe the quality of a paper should not be assessed exclusively by its ultimate theoretical success, but also by whether it raises interesting questions and approaches them in a novel yet rigorous way.

What distinguished Goodman from the other early writers and earned his article a prominent place in the debates on this topic to this day, was how squarely he confronted the problem of truth conditions of counterfactuals. He briefly addressed antecedent ambiguity, but he rightly observed that a solution to this problem could not be overstretched to also cover the problem of restrictions on the set of relevant conditions, which remains even if ambiguities are resolved and ties are broken.[22]

In order to approximate the intuitive truth conditions of counterfactuals, Goodman, argued that a number of requirements must be placed on the set of true relevant conditions S in the argument underlying the subjunctive $A \boxdot\!\!\rightarrow C$. I shall enunciate a groomed version of these conditions, as they emerged from Goodman's discussion with Parry[23], trying to iron out some technical imprecisions that somewhat obscured the final analysis.

1. $\{A\} \cup S$ must be both logically and non-logically consistent, or 'self-compatible'. By 'non-logical self-incompatibility' Goodman means violation of a non-logical (e.g. causal) law. An incompatible set of premises would of course render any underlying argument valid, irrespective of the consequent. Non-logical incompatibility allows this by dint of empty covering laws.

[22](Goodman, 1947), pp.115-116
[23]See (Parry, 1957) and (Goodman, 1957).

2. $\neg A$ must be compatible with each $S_i \in S$, that is, the denial of the antecedent must not strictly imply any relevant condition in S, nor may any S_i contain as an essential component a proposition strictly implied by $\neg A$. This rather verbose requirement is introduced by Parry in order to get rid of conditions that strictly depend on the fact that the antecedent is false and therefore do not deserve to be retained in counterfactual scenarios. A trivial example is $\neg A \vee C$, which in case of a counterfactual will be true regardless of C, thus validating any underlying argument.[24]

3. S must be compatible both with C and $\neg C$. That is, S alone should not decide about C—only after adding A to S should everything fall into place. Compatibility of S with $\neg C$ guarantees that $A \boxdot\!\!\rightarrow C$ will not come out as true only because C happens to be true (typically in conditionals of the 'even-if' sort).[25]

4. $\{A\} \cup S$ must lead by law (again, logical or non-logical) to C. This guarantees that the set S really does the trick of validating the underlying argument.

5. There must be no alternative set of truths S', compatible with C[26], such that $\{A\} \cup S'$ is self-compatible and leads by law to $\neg C$. This must be required as we are never prepared to accept both $A \boxdot\!\!\rightarrow C$ and $A \boxdot\!\!\rightarrow \neg C$ as true. The idea is that if there are two alternative sets of facts satisfying the other conditions and leading respectively to C and $\neg C$, then we are bound to reject both subjunctives.

These requirements make up quite a mouthful, but seem to square reasonably with our pre-theoretical use of subjunctives. But are they also sufficient? Despite all the efforts, they can be shown to be incomplete.

Take first an 'even-if' conditional such as

(49) If Jones were in Carolina, he would still be a moron.

Here the subjunctive might be true solely in virtue of its consequent happening to be true and the antecedent incapable of changing anything about that fact. But according to requirement 3 put forward by Goodman, S has to be compatible with both C and $\neg C$, so we cannot simply include 'John

[24]The rider ensures, in quite a radical way, that such inadequate facts don't sneak in as parts of more complex propositions.

[25]Requiring compatibility of S with C is actually redundant in view of the other requirements, but following Goodman, I state it for perspicuity's sake.

[26]The rationale for this provision is the same as in the case 3.

is a moron' in our collateral premises. However, if we forgo requirement 3, we will get too many of such subjunctives:

(50) If Jones were on the North Pole, he would still be wearing a suit and a bow tie.

You might argue that (50) is still excluded by our other requirements—not because S is incompatible with $\neg C$, but rather because S is incompatible with A (say, in virtue of a 'law' that stipulates that no-one in their sane mind wears a suit and a bow tie on the North Pole). However, general recourse to such an approach clashes with non-monotonicity of subjunctives with respect of the strength of their antecedents. In most cases we are capable of coming up with possible additional circumstances that would render a counterfactual more palatable:

(51) If Jones were on the North Pole shooting a TV stunt, he would be wearing a suit and a bow tie.

A seems to have been compatible with S after all.

Our predicament does not end here, however. As Goodman himself points out, more should be excluded from S, as it might contain propositions that *wouldn't be true if the antecedent were.* Take (42), for example

(42) If the match had been struck, it would have lighted.

The match wasn't struck and it didn't light. Now if you include the fact that it didn't light in your S, based on the very same law as in the latter case you may end up with a counterfactual such as

(42*) If the match had been struck, it would not have been dry.

The reason being that as much as in (42) the law excluded the possibility of not lighting of a dry match struck in the presence of oxygen, in (42*) the law excluded the possibility of being dry of a non-lighting match struck in the presence of oxygen.

However, even though (as befit them) the members of the liberal wing didn't find (42*) objectionable, for most of us it clearly isn't a patch on (42). There seems to be a fundamental asymmetry between the condition of dryness and lighting that our analysis fails to account for. Dryness, as opposed to non-lighting, *is* transposed to the counterfactual situation introduced by striking the match.

Unlike with the other problems, the structure of this puzzle doesn't depend on the particulars of Goodman's approach, so that any account of

counterfactuals that relies heavily on covering laws will have to cope with it. We shall see that most non-trivial such theories will struggle to explain the asymmetry between matters of fact that this problem reveals and as a result many authors will choose to posit this asymmetry by fiat.[27] Because of its importance, this problem is often referred to as *Goodman's puzzle* in the literature.

In an attempt to address the difficulty, Goodman's final analysis replaces the notion of compatibility in the requirements 1-5 with the notion of *cotenability*:

Cotenability A is cotenable with S if it is not the case that S would not be true if A were true.[28]

While this analysis is not completely trivial[29], recourse to cotenability, of course, leaves it blatantly circular. And although there are cases of virtuous circularity in semantic analysis, this doesn't seem to be one of them. The counterfactual in the definition of cotenability supports too much weight of the analysis for it to retain any interest. The task of explaining how counterfactual scenarios are constructed, while avoiding the use of subjunctives for this purpose, has haunted writers on this subject ever since. As we will see, technical notions such as 'similarity of possible worlds' (Stalnaker & Lewis) or 'lumps of thought' (A. Kratzer) have been introduced, with varying degrees of success, for the very purpose of warding off this kind of circularity.

[27] See in particular Section 5.3.

[28] Goodman defines self-cotenability of the set $\{A\} \cup S$ as cotenability of A with S, which poses some problems—since $A \boxrightarrow C$ doesn't in general imply $C \boxrightarrow A$, self-cotenability of sets of propositions doesn't seem to be well-defined. Also, as a side-remark, it is worth mentioning that self-cotenability is a strictly stronger notion than that of self-compatibility, which it replaces.

[29] For instance, it validates Conditional Excluded Middle as follows: take any true C. Then if it is not the case that $A \boxrightarrow \neg C$, by Goodman's definition of cotenability, $\{C\}$ is an acceptable value for S and so trivially $A \boxrightarrow C$. Hence in any case, $(A \boxrightarrow \neg C) \vee (A \boxrightarrow C)$. For a false C, $\neg C$ is true and the conclusion follows by the same argument and Double Negation. Therefore, if CEM is to be rejected for subjunctives (as many have argued that it should), Goodman's condition in the definition of cotenability of S with A must be strengthened to something closely resembling $A \boxrightarrow S$, making the circularity even more obvious. This is the general notion of cotenability that I shall employ for the purposes of comparing cotenability approaches with possible world ones—see Section 4.5.

3.2. Laws

We have seen that metalinguistic theories distinguish between actual matters of fact relevant for the truth of subjunctives (relevant conditions) and general connecting principles or lawlike generalisations that sustain the inference from the relevant conditions and the antecedent to the consequent. So even if we knew how to determine relevant conditions, we would still need to provide criteria to tell lawlike generalisations from the merely accidental ones in order to provide a satisfactory and complete analysis of subjunctives. As Goodman showed, accidental generalisations are not up to the task of shoring them up. Whereas the fact that in vacuum all objects fall with equal acceleration supports

(52) If this hammer and this feather were dropped from the same height in the vacuum, they would reach the ground simultaneously,

the accidental fact that all coins in my pocket are silver doesn't support

(53) If this copper coin were in my pocket, it would be silver.

As with relevant conditions, we find precious little consensus on how to account for this difference in the early literature.

Chisholm's liberalism accorded no special position to universal generalisations, whether accidental or lawlike. They could be scraped along with any other propositions the speaker deemed undesirable. That's how you could assert a counterfactual such as (44), which contradicts an arguably non-accidental generalisation 'All men are mortal'.[30]

Rescher, as I have remarked, roughly shared Chisholm's *laissez-faire* approach, but among collateral premises he gave pride of place to covering laws. There is no further explanation of the lawlike nature of some universal generalisations—it is not that they sustain subjunctives because of being nomological, but they are nomological because we use them to sustain subjunctives:

> [...] because the nomological use of counterfactuals represents a determination to retain the appropriate covering generalization [...] at the cost of adapting all else to it.
>
> (Rescher, 1961), p.190

Naturally, this does not really leave us much better off, since the question as to why we give preference to certain generalisations rather than to others when we assess subjunctives remains unanswered.

[30](Chisholm, 1946), p.303.

Mackie had more interesting things to say about laws. First, he remarked that nomological/accidental character of a generalisation was not codified semantically. The same generalisation can be treated as merely accidental or lawlike without inducing any change in the universal proposition itself. Where the difference lies, according to Mackie, is in the sort of grounds on which we believe the generalisation in question. To be justified in treating a universal as a lawlike generalisation means to believe it on good inductive grounds—that is why a causal law enables us to 'go beyond the cases for which the law has been checked, and so to advance open conditionals', as well as counterfactuals, whose antecedents are prone to alter the extension of the subject term.[31]

However, the difficulties with such an account of lawlike generalisations had already been pointed out a whole thirty years earlier by Frank P. Ramsey:

> The difference according to Braithwaite is that universals of law are believed on grounds which are not demonstrative. This will not do because
>
> (a) some universals of law are not believed at all
>
> (b) some universals of fact are believed on non-demonstrative grounds
>
> (c) some universals of law are believed on grounds which in his sense are demonstrative.
>
> (Ramsey, 1928), p.128

As for (a), Ramsey argues that to say that some characteristics of the offspring are caused by certain unknown genes does not mean to say that there is good inductive evidence for the link between some genes and the characteristics in question. The purpose of his example is to show that the idea of a causal relationship between two phenomena does not seem necessarily to concern the grounds upon which it is ascertained—we know that a certain property must have been caused by a gene, even though we do not know which particular one, and so cannot have any inductive evidence about this relationship. Braithwaite and Mackie might perhaps argue that this quantification involves misleading scope ambiguity, but at a second glance, this looks implausible. Surely there is a clear difference between the causal law that ensures that there are genes causing certain phenotypic characteristics and the causal law between concrete genes and

[31] (Mackie, 1962), p.73. Mackie calls 'open conditional' an indicative conditional whose antecedent remains undecided, that is, a typical specimen of this kind of conditional.

the characteristics in hand. Nothing in Ramsey's example prevents us from giving the existential quantifier a large scope.

Point (b) is also relevant for our discussion. Ramsey points out that certain accidental generalisations are believed on grounds that are not demonstrative, such as testimony or some kind of *reductio*. But they can undoubtedly also be believed on inductive grounds: imagine I run into a sizeable anti-government protest in the centre of the city. I cannot oversee all the participants, but by the insignia, banners and chants of those in my immediate surroundings I can make out that they are socialists. I conclude by an inductive argument that all participants in the demonstration are socialists. But this, of course, remains an accidental generalisation, incapable of sustaining a counterfactual such as

(54) If I took part in the protest, I would be a socialist.

There may of course be good reasons to oppose government policy while being liberal or even communist.

As regards (c), Ramsey admits it is not as clear as the other points and his supporting example indeed doesn't strike me as very convincing:

> 'Whenever this balloon was filled with hydrogen and let go, it rose'; this, or something like it only more complicated, is surely (*sic*) a universal of law and yet might be believed as a result of observing all its instances.
>
> (Ramsey, 1928), p.129

I doubt whether nomological generalisations are ever believed on grounds that are exclusively demonstrative (that is, based entirely on an inspection of all their instances, as opposed to using these instances in an inductive inference to an underlying causal link). This is in all likelihood a conceptual problem. Someone who believes that she has proven a nomological generalisation demonstratively (in Ramsey's sense) can be construed simply as not having grasped the difference between accidental and non-accidental generalisations. On the other hand, checking the very same instances and accepting a causal law on their strength may be quite unexceptionable, as long as it is not suggested that checking the instances is 'all there is to justification' of such a law. Checking instances is never by itself completely sufficient to warrant belief in a causal law.

It was his own account of universals of law that forced Ramsey to accept something like (c). According to Ramsey, the special nature of these universals was due to their aptness as fundamental generalisations (that is, general axioms and theorems deduced without recourse to existential statements) in the system of all knowledge. This is a very peculiar theory, which of course has to explain how we can bestow such a status on a generalisation even if we, as a matter of fact, don't know everything. According to Ramsey, we organise what we do know in a deductive system and we ask ourselves *what the effects on the system would be if we knew a little more.* The axioms and general theorems of the resulting system are then what we call 'laws'.[32]

This theory was adopted by Lewis, who used it to explain why criteria of similarity between possible worlds placed such an extraordinary emphasis on laws.[33] But Lewis's approach purports to avoid Ramsey's reliance on omniscience, shifting the emphasis to deductive systems as abstract objects instead. Some of the latter are true, and from among them we can select those that best combine perspicuousness and logical strength. Laws of nature are those universal generalisations that feature as axioms or theorems in all the best performing systems.

Some modification seems indeed necessary in order to give plausibility to Ramsey's theory, as falling back on omniscience is clearly ineffectual.

[32](Ramsey, 1928), p.131
[33](Lewis, 1973a), p.73

We can of course entertain the counterfactual possibility of knowing all there is to know, but such a mental experiment leaves us in the dark about what our beliefs then would be (otherwise the possibility wouldn't be counterfactual anymore and we would know everything already as it is!). But it is by no means clear whether Lewis's account fares any better on this point. For how could we ever be justified in calling any generalisation a natural law, given the widely acknowledged limitations of our knowledge of the world?

Lewis contends that his take on the problem doesn't require us to envisage the best axiomatisations themselves—it is enough that we try to approximate this ideal to the best of our capacities. But then the recourse to the best possible deductive system seems utterly gratuitous. The generalisations that we treat as laws will simply turn out to be those propositions appearing in *our* best deductive systems, whose main purpose is to systematise the knowledge we have (as opposed to the knowledge we would have if we knew everything). So, all things considered, for Lewis the difference between laws and accidental generalisations boils down to a matter of scope, a law being a generalisation that covers vast swathes of fact in one fell swoop.

Ramsey very soon realised that his idea didn't work, but instead of settling for a conservative revision along Lewis's lines, he opted for a radical change of picture. The purpose of laws does not reside only in systematising the welter of known facts, as might seem reasonable from a semi-divine vantage point granted by possible-worlds semantics—in real life a law often precedes the facts it is meant to govern. Lewis's approach is very neat at a high level of abstraction, but it faces severe difficulties explaining the role of laws in our mental lives, which are fraught with uncertainty.

Thus, a mere year after his first take on the problem of the distinction between accidental and lawlike generalisations, Ramsey returned to it in another pregnant paper, calling the former *conjunctions* and the latter *variable hypotheticals*. Variable hypotheticals, according to the new theory, express 'an inference we are at any time prepared to make, not a belief of the primary sort.'[34] This time round, a sharp line of divide separates propositions, which allow us to get our bearings in the physical world, and variable hypotheticals, or causal laws, which set forth rules for inferring propositions:

> Variable hypotheticals are not judgments but rules for judging 'If I meet a ϕ, I shall regard it as a ψ'. This cannot be *negated* but it can be

[34](Ramsey, 1929), p.134

> *disagreed* with by one who does not adopt it.
>
> (Ramsey, 1929), p.137

The interesting and confusing point about variable hypotheticals is that if we try to enunciate their truth conditions or apply them in practice, we end up with mere material conditionals, which have the status of conjunctions. But this does not mean that our distinction between variable hypotheticals and conjunctions is spurious, but rather that the import of the former cannot be fully expressed in terms of truth conditions. Variable hypotheticals, on Ramsey's account, sustain both future predictions and 'assertions about unfulfilled conditions'[35], that is, counterfactual conditionals, providing an interesting and much neglected link between indicatives and subjunctives.

Most of these points were explicitly acknowledged by Goodman, even though he didn't sever laws from ordinary propositions as radically as Ramsey did. In order to exclude hybrid cases, he defined non-accidental generalisations as those statements whose 'acceptance does not depend upon the determination of any given instance.'[36] But he also set out to inquire what the rules of acceptance of laws, as general statements that go beyond particular instances, *should* be, as he found most contemporary theories of confirmation grossly inadequate for the task. In an attempt at a solution, he put forward a non-formal theory of induction and suggested that a hierarchy of universals statements, ranked by their degree of entrenchment, could account for differences in their aptness for sustaining counterfactuals.[37] However, it is by no means clear whether this notion of entrenchment can be satisfactorily defined in terms of confirmation—it is not the certainty with which I or anyone might believe that all the coins in my pocket are silver that impinges on the acceptability of (53), but rather something like its inferential role within the system of our beliefs. One of our goals should therefore be to clarify the latter notion.

Be it as it may, if we accept that there is more than a grain of truth to Ramsey's and Goodman's treatment of counterfactuals, we will still need to develop their proposals further in view of the plain fact that apparently *by far not all subjunctives are sustained by lawlike regularities*. Take, for example, Edgington's example

(55) If I had known you were coming, I'd have baked a cake.[38]

[35](Ramsey, 1929),p.143
[36](Goodman, 1947), p.126
[37](Goodman, 1983), p.122
[38](Edgington, 2007), p.140

You might come up with a pseudo-law underwriting this subjunctive in order to save the day for metalinguistic theories, but it surely seems as a highly artificial move. We can easily imagine a follow-up dialogue between the original speaker, her niece and a by-stander:

You: Oh, so whenever you know she's coming, you bake her a cake?

Dorothy: Well... not always, but this time I had plenty of flour and chocolate in the pantry.

You: Ok, so whenever you know she's coming and you have all what you need, you bake her a cake.

Dorothy: Usually I do, but I must feel like it, of course. And then, I don't even need to have all the ingredients ready—if I have time, I can always do some shopping.

Dorothy's niece: But don't forget that if I'm currently on a diet, you make a salad instead.

The quest for all the relevant features of the situation that shore up the counterfactual in (55) can go on for quite some time, and even then remain inconclusive. Full generality is notoriously difficult to come by when non-accidental generalisations are involved and even our best laws contain a tacit *ceteris paribus* clause. We tend to view this as a nuisance and therefore dismiss it by chalking it up to the existence of myriads of theoretically possible, but practically irrelevant defeating conditions for any given law. However, the ubiquity of counterfactual reasoning, which more often that not relies just on such defeasible regularities, reminds us that this is the default human way of approaching generality and should therefore deserve a closer and more equitable inspection.

Edgington puts her finger on a signal issue—our use of counterfactuals precedes the highly theoretical activity of formulating natural laws. The inferences which they make manifest more often than not fall short of the lofty standards required of exact sciences; on the contrary, they tend to be highly contextual and seldom readily systematisable. Let us for a moment return to Goodman's original example:

(42) If the match had been struck, it would have lighted.

I can easily imagine my grandmother assenting to or even asserting this counterfactual, while being quite sure she has never given a moment's thought to the alleged natural law sustaining it and that she probably does

not know that the presence of oxygen is necessary for the process to take place (though she does take good care to store her matches in a dry place). Now, of course, linguistic meaning being determined by social norms, a speaker does not always have complete control over the truth conditions of his assertions. But even if the rendering of truth conditions of subjunctives in terms of something like the Basic Schema were roughly right, we would still miss an account of what people unaware of the elements of such an analysis actually try to accomplish by asserting them. As (55) and the urge to make the underlying inference explicit by means of something like a variable hypothetical suggest, cotenability theories, with their focus on logically sound arguments, seem to overly intellectualise the practice of counterfactual reasoning. If we really wish to get to the heart of the matter, I submit that we must treat my grandmother as a competent user of counterfactuals, rather than a mere dilettante.

3.3. Conclusion

We have seen that a review of the early literature on subjunctives raises many an interesting problem, which the later authors would have to face in one way or another. It offers, therefore, a good vantage point to assess the conceptual innovations of their successors. Moreover, the approach of metalinguistic theories, while perhaps excessively rationalistic, keeps reasonably close to the actual uses of subjunctives (at least in some domains), with a view to making them comprehensible. By analysing newer proposals, which sometimes tend to dodge this desideratum, in the light of their forebears, we can reasonably hope to re-establish the missing link between the practical use of conditionals and theoretical accounts that strive to explain it.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 4

Similarity

Semantic considerations have accompanied modern modal logic since its very beginnings, as the following passage by C. I. Lewis attests:

> Any set of mutually consistent propositions may be said to define a "possible situation" or "case" or "state of affairs". And a proposition may be "true" of more than one such possible situation—may belong to more than one such set. Whoever understands "possible situation" thereby understands "consistent propositions", and vice versa. [...] In these terms, we can translate $p \dashv q$ by "Any situation in which p should be true and q false is impossible".
>
> (Lewis, 1918), p.333

This opens the theoretical possibility of evaluating propositions in different possible situations, construed as consistent sets of propositions. But what do we mean by consistency anyway?

Two principal notions of consistency have developed since the dawn of classical logic: the logical or model-theoretical notion on the one hand, and the logically stronger semantic or modal one on the other. While '*a* is red' and '*a* is blue' are intuitively and therefore modally inconsistent, they are consistent on the logical conception, as there is no way to derive a contradiction from two propositions devoid of logical operators.[1]

For this reason, early logicians assumed that sets comprised exclusively of atomic propositions were always consistent—inconsistency was always put down to the presence of logical operators. This attitude was reflected in the picture of language put forward in Wittgenstein's *Tractatus* and later

[1] By the completeness theorem for first-order logic, any two simple propositions must be simultaneously satisfiable. That is so, because an interpretation does not have to attend to the actual meaning of predicates when fixing their extension.

adopted by many logical positivists. Nevertheless, it did not take long for them to realise that there was a stronger notion of consistency at work in natural language. Later theorists made use of this latter notion to underpin the idea of a possible world and reinterpreted Lewis's strict implication accordingly. This is the convention I will follow in this essay as well. Thus, 'He is a bachelor' strictly, though not logically, implies 'He is unmarried'.

However, C.I. Lewis's work in logic was still firmly anchored in the axiomatic-deductive tradition of *Principia Mathematica* and had not been affected by the semantic revolution ushered in by Gödel's and Tarski's work in the 1930s. Attempts to formulate a model theory for modal logic began with Carnap, who used sets of atomic sentences (which, as we have seen, were by definition deemed logically consistent) as elements of the universe of his model. The two crucial improvements introduced by subsequent modal logicians concerned the possibility of relaxing the accessibility relation on the universe (which Carnap fixed as universal), and the use of the full-fledged notion of possible world, which supplied a complete valuation of the set of atomic sentences of a language (and thus of their Boolean compounds as well).[2]

In the 1970s, the semantic turn in modal logic reached also the literature on counterfactuals. Theories based on the notion of possible world, drawing heavily on Kripke's legacy, replaced the earlier metalinguistic approaches and have come to dominate the literature on counterfactuals. Pioneered by R. Stalnaker and D. Lewis, they added another parameter to normal modal models in order to enable the valuation of counterfactuals at possible worlds: a selection function (Stalnaker) or an ordering relation based on relative similarity (Lewis). One of the advantages of such an approach is that particular choices in semantics (for instance, imposing different constraints on the ordering relation) come hand in hand with particular outcomes in resulting logic, which can in turn be tested against actual language use. We will therefore first examine the formal structure of these theories and only then will we address the related philosophical and linguistic issues.

[2]In particular, this involves work by A. Prior, S. Kanger, J. Hintikka and especially S. Kripke from the 1950s and early 1960s. See (Korte et al., 2009), p.536.

4.1. The basic framework: Stalnaker

4.1.1. Semantics

Although Stalnaker's theory can be considered as a particular case of Lewis's, for historical and expositional reasons I will present the former first and only then analyse the rationale behind the generalisations put forward by Lewis. Stalnaker's theory aims at providing a general semantic framework for all conditionals, both indicative and subjunctive, while Lewis generalised Stalnaker's theory for subjunctives, but espoused F. Jackson's account of indicatives.

Stalnaker's semantics for conditionals comprises two basic components:[3]

1. *A model structure*, that is, a modal logic model $\mathcal{M}$. It equals an ordered triple $\langle K, R, \lambda \rangle$, where K is the universe of all possible worlds, R is an accessibility relation that permits the valuation of formulae prefixed by standard modal operators $\Box$ and $\Diamond$, and λ is a special impossible possible world, designed to deal with impossible antecedents. Except for λ, $\mathcal{M}$ is a standard model for a normal (that is, Kripke-frame) modal logic.[4]

2. *A world-selection function* $f : \mathcal{P}(K) \times K \rightarrow K$. That is, f takes a proposition and a possible world into a possible world. The world-selection function is interpreted as an abstraction from a 'change function' which represents the subject's methodological policies for belief revision by taking an acceptance state (set of possible worlds) and a proposition into another acceptance state. For Stalnaker, conditional propositions are the product of projecting such policies onto the world, so they should be interpreted according to a change function of someone who knows all the facts.[5] The world-selection function f, reducing acceptance states to just one actual world, aims to accomplish just that.

[3](Stalnaker, 1968), p.45

[4]Actually, in a classic modal semantics, the model explicitly contains a valuation function V, which assigns a truth value in a possible world to any wff of the language (See (Blackburn et al., 2004)). This allows a recursive definition of truth for formulae of any degree of complexity. However, as this is not one of our goals at the present moment, we will follow Stalnaker in treating the valuation as implicit in the definition of a possible world. Henceforth we shall also reserve small Greek letters for possible-world variables.

[5](Stalnaker, 1984),p.120

Neither R. Stalnaker nor D. Lewis followed C.I.Lewis's preliminary identification of possible situations with sets of consistent sentences. They both contended that our capacity for distinguishing between relevant alternative ways the world might be was prior to their linguistic articulation. In this section we will therefore reverse the order of explanation and conceive propositions, as has become customary within possible-world semantics, as subsets of possible worlds.[6] The price is that possible worlds will be treated as unanalysed objects and propositions as their collections.[7] While this is technically convenient, it doesn't commit us one way or other as far as the status of possible worlds is concerned—this issue will be taken up in Section 4.3. As a side-effect, we will also have to live with a separate treatment of λ, which will be in the range of f, but will not appear in any member of its domain (that is, the impossible proposition will be treated as $\emptyset$ rather than $\{\lambda\}$).

In this model, each possible world α will be associated with the set of possible worlds accessible from it by way of the relation R; in set-theoretical notation this set is rendered as $[\alpha]_R$. We can think of $[\alpha]_R$ as the outer limits of what is imaginable or otherwise relevant for evaluations of conditionals in α—nothing beyond this set has any bearing on their truth values, even if the only A-worlds in K were in $K - [\alpha]_R$. It will therefore be useful to define restrictions of propositions to these sets. Let us therefore define $[\![A]\!]_\alpha$ (in $\langle K, R, \lambda\rangle$) as $[\![A]\!] \cap [\alpha]_R$.

With these ingredients in hand, we can enunciate the truth conditions of conditionals in a Stalnaker system:

TC Conditionals Stalnaker
$[\![A > C]\!]^{\mathcal{M}} = \{\alpha \in K \mid [\![A]\!]_\alpha = \emptyset \vee f([\![A]\!], \alpha) \in [\![C]\!]\}$

In other words, a conditional $A > C$ is true in a possible world of $\mathcal{M}$ if there are no accessible antecedent worlds or if the antecedent A forces the selection function to switch to a world where C is true. Stalnaker's intent is for f to reproduce on an ontological level the procedure of adjusting the system of one's beliefs in view of an entertained hypothesis, envisaged by the Ramsey test. This ontological turn is needed, according to Stalnaker, in order to firmly anchor the semantics of conditionals (Ramsey test provides

[6] This will also allow us to replace truth talk with the more economical set-theoretic notation—instead of 'A is T in α' we can simply write '$\alpha \in [\![A]\!]$', where $[\![A]\!]$ is defined as the set of possible worlds satisfying A, that is, as the proposition that A.

[7] In standard modal logic, possible worlds can even be viewed as elements of a first-order model. That's why we will be able to spell out the truth conditions of counterfactuals in a first-order language.

a decision method for evaluating the truth of $A > C$, but not its actual objective truth conditions).[8] For this idea to work properly, we are compelled to impose the following constraints on f:

1. For all α and A such that $[\![A]\!]_\alpha \neq \emptyset$, $f([\![A]\!], \alpha) \in [\![A]\!]$. If $f([\![A]\!], \alpha)$ is to represent the result of the adjustments made to a system of beliefs corresponding to α in order to accommodate A, at the very least it has to support A. This also guarantees conditionals of the form $A > A$ to be evaluated as true in any world of any model (in other words, they will be logically true).

2. For all A and α, $f([\![A]\!], \alpha) = \lambda$ only if $[\![A]\!]_\alpha = \emptyset$, or in other words, the selection function will output λ, the impossible world, only if A is inconsistent from the point of view of α. This ensures that f is defined on the whole $\mathcal{P}(K)$.[9]

These constraints on the selection function are arguably minimal. However, together with the stipulation that f has to be a function, they set in place the machinery capable of warding off the classical paradoxes of material conditional, while respecting some of the strongest intuitions about what a formal theory of conditionals should aim for. Neither false antecedents nor true consequents make Stalnaker's conditionals trivially true, which makes this treatment viable also for counterfactuals. Stalnaker's original idea was, of course, to put forward an Y-shaped account of conditionals, explaining the difference between indicatives and subjunctives by an epistemic restriction on K; this, however, proved more difficult than expected.[10]

The two further constraints are devised to reflect the intuition that the belief revision mandated by the Ramsey test should in some sense be minimal. This implies that the resulting belief system should not deviate gratuitously from the original one—the only permissible modifications are those necessary to accommodate the hypothesis expressed in the antecedent. Hence the notion of *similarity* or *resemblance* to the actual world,

[8] (Stalnaker, 1968), pp.44-45

[9] In order to be able to treat antecedents as set of possible worlds, I have had to adjust Stalnaker's constraints slightly. Stalnaker, who doesn't represent impossible antecedents as the empty set, can let Constraint 1 hold unrestrictedly, stipulating that all propositions are true in λ, which together with Constraint 2 makes $A > A$ logically true also for inconsistent As. However, that would in our case lead to $f(\emptyset, \alpha) \in \emptyset$, which conceals a contradiction. We obtain the same result by explicitly allowing for the trivial case in our truth conditions. That entails, nonetheless, that λ only plays an inessential role in our system.

[10] See Section 4.1.3.

which is to inform the selection function by providing an ordering of the possible worlds eligible as its values—given any A and α, f is set to output the least β under $\preceq_\alpha$ such that $\beta \in [\![A]\!]$.

3. For all A and α, if $\alpha \in [\![A]\!]$, then $f([\![A]\!], \alpha) = \alpha$. If a system of beliefs already sustains a hypothesis, its addition shouldn't affect the system in any way. This makes α the $\preceq_\alpha$-least element on the whole field of the relation.

4. For all α and antecedents A and A', if $f([\![A']\!], \alpha) \in [\![A]\!]$ and $f([\![A]\!], \alpha) \in [\![A']\!]$, then $f([\![A']\!], \alpha) = f([\![A]\!], \alpha)$. This guarantees that the distance of two worlds with respect to the actual one will not be antecedent-relative. Together with the fact that f is a function, it also ensures that the underlying relation be transitive and well-founded on K.

We can now reveal the relation on K defined by f and survey its properties. Given any α, let us define $\preceq_\alpha$ as follows: $\beta \preceq_\alpha \gamma$ iff $f(\{\beta, \gamma\}, \alpha) = \beta$.[11] It is easy to show that $\preceq_\alpha$, as defined, is a *non-strict well-order*.[12]

4.1.2. Logic

The constraints we have discussed so far are, of course, not nearly sufficient to yield a unique s-function (how Stalnaker addresses this issue will be the topic of the following section), but they do shape a conditional logic, which he calls C2. The strict constraints on f (especially Constraint 4) in turn make C2 very strong. Since C2 is the strongest system of logic that I will consider in this dissertation, it will be convenient to spell out its axioms in full for later reference:[13]

Stalnaker's Logic C2

[11] Here is why this is reasonable: for any antecedent A, f is meant to select the world most similar to α that satisfies A. So if for the antecedent $\{\beta, \gamma\}$ f selects β rather than γ, it must be because β resembles α more than γ does.

[12] See A.1 for details of the proof.

[13] In doing so, I will closely follow (Veltman, 2012) and (Veltman, 1985). Veltman's presentation, as opposed to Stalnaker's, avoids the use of any defined notions apart from the conditional connective $>$. The initialisms stand, respectively, for: Tautology, Modus Ponens, Replacement of Equivalents (in the antecedent/consequent), Conditional Identity, Conjunction of Consequents, Consequent Weakening, Antecedent strengthening by consequent, Disjunction of Antecedents, Antecedent Strengthening with Possibility, Modus Ponens for $>$, Conjunctive Sufficiency and Conditional Excluded Middle.

(Taut)	Any classical tautology is an axiom.
(MP)	If A and $A \supset B$ are theorems, then B is a theorem.
(RE)	If $A \equiv B$ is a theorem, then
	a) $A > C \equiv B > C$ is a theorem
	c) $C > A \equiv C > B$ is a theorem
(CI)	$A > A$
(CC)	$((A > B) \wedge (A > C)) \supset (A > (B \wedge C))$
(CW)	$(A > B) \supset (A > (B \vee C))$
(ASC)	$((A > B) \wedge (A > C)) \supset ((A \wedge B) > C)$
(AD)	$((A > C) \wedge (B > C)) \supset ((A \vee B) > C)$
(ASP)	$(\neg(A > \neg B) \wedge (A > C)) \supset ((A \wedge B) > C)$
(MP$^{>}$)	$(A > C) \supset (A \supset C)$
(CS)	$(A \wedge C) \supset (A > C)$
(CEM)	$(A > C) \vee (A > \neg C)$

In what follows, we will examine if a case can be made for weaker systems. The relation of comparative similarity $\leq_\alpha$ provides a convenient handle on such weakening—a *well-order* being the strongest ordering relation at our disposal, we can check if switching to a weaker one would do more justice to our intuitions.

However, before doing so, we should have a look at some important inference patterns rendered invalid by C2, and *a fortiori* by all weaker similarity-based systems.[14]

Antecedent strengthening: $A > C \nvdash (A \wedge B) > C$
Stalnaker-Lewis systems are built up explicitly with the purpose of steering clear of this pattern, the invalidity of which sets natural language conditional reasoning apart from the monotonic conditional reasoning proper to deductive systems. They manage to avoid it by allowing for the possibility of the closest $A \wedge B$-world(s) being further away from the actual world than the closest A-world(s), which opens the door to different valuations of C in each of them.

Transitivity: $A > B, B > C \nvdash A > C$
Transitivity can be regarded as a peculiar kind of antecedent strengthening, and as such it is equally liable to criticism. When A is far enough, it might happen that even though B counterfactually implies C, A fails to do so.

Contraposition: $A > C \nvdash \neg C > \neg A$

[14]For detailed proofs, see A.1.

Practically all counterexamples to contraposition in natural language involve 'even-if' conditionals.[15] It may so happen that in an acceptable conditional 'if A, then still C', A is a necessary, but singly insufficient condition for $\neg C$. In such case, of course, $\neg C > \neg A$ will be unacceptable.

Simplification of Disjunctive Antecedents: $(A \vee B) > C \nvdash (A > C) \wedge (B > C)$
While all the above examples were of inference patterns that we were keen to dismiss anyway, this case is different. In natural language, we tend to reason along SDA's lines both with indicatives and subjunctives. However, it is very easy to falsify SDA on a Stalnaker frame—it suffices to guarantee that whichever of the two disjuncts is further away from actuality doesn't, as opposed to the one closer by, counterfactually support the consequent.

Adding SDA to C2 is a non-starter, as the resulting system would validate antecedent strengthening and thus convert $A > B$ into a strict conditional:[16]

1. $A > C$	Premise
2. $A \equiv (A \vee (A \wedge B))$	By Taut
3. $(A \vee (A \wedge B)) > C$	From 1,2 by REa)
4. $(A > C) \wedge ((A \wedge B) > C)$	From 3 by SDA
5. $(A \wedge B) > C$	From 4 by propositional logic

⊣

Observe that Replacement of Equivalent Antecedents is the only potentially controversial rule of inference actually used in the proof—as most similarity-based systems accept it, we will not be able to enrich any of them with SDA without trivialising the resulting logic. This is also the reason why most authors keen on keeping SDA on board, similarity theorists or otherwise, will be compelled to reject REa).

An exception is (Nute, 1975), where REa) is effectively quashed. However, this strategy leads to new difficulties—how should we keep the many legitimate uses of REa) apart from the offending ones? And how should the counterexamples against SDA be accounted for? In both cases, Nute appears to draw *ad hoc* distinctions, either between acceptable and unacceptable replacements of logical equivalents, or between allegedly different kinds of conditionals.[17]

[15] See (Lycan, 2001),p.34

[16] In what follows, I shall simplify the proofs by omitting trivial uses of MP, in fact treating most axioms as natural deduction inference rules.

[17] See (Nute and Cross, 2001), p.31, where he retracts his previous position for these very reasons.

But wait, this semantic quagmire seems to be just the perfect occasion to use our marvellous panoply of Gricean remedies. Banishing SDA from our semantics, while accounting for it on pragmatic grounds would do the trick—we could eat our cake and keep it at the same time. This is the view endorsed by Bennett[18], as well as some others. However, similarity-based approaches face serious difficulties explaining the intuitive validity of SDA by means of conversational implicature. Granted, asserting $(A \vee B) > C$ on the basis that $A > C$ while regarding B as remote and rejecting $B > C$ is always misleading, if only because you could make your point both more simply and accurately by asserting $A > C$. However, why the audience should infer from your use of a disjunctive antecedent that you accept both simplified counterfactuals never mind the remoteness of their antecedents is by no means clear. You might rather be interpreted as signalling that you do not know which one of the antecedents is closer to actuality, or that each of them is as close as the other. Alonso-Ovalle has shown that these difficulties are compounded with Lewis's interpretation of might conditionals in the absence of uniqueness assumption.[19] Furthermore, as Lycan pointed out, if SDA's apparent validity is due to a conversational implicature, it should be cancellable.[20] Bennett says that to find a pertinent cancelling context is a matter of a 'five-finger exercise' but, curiously, he doesn't find enough space to spell it out and check whether it is convincing.[21]

Another option would be to relinquish the boolean semantics of 'or' and, for subjunctives, slightly adapt the workings of the 'would' modality. This is quite a radical step that not everyone is happy to make—it introduces an alien element into our otherwise neat and well-behaved logical system. Lewis himself admitted to qualms about tinkering with it, but despite the reluctance, he conceded that there was a strong case for a non-boolean semantics for 'or', even though he did not undertake such a project himself.[22] Alonso-Ovalle also claims an independent motivation for this treatment of SDA in his (Alonso-Ovalle, 2009), building upon the idea that natural language 'or' supplies a set of semantic alternatives, the interpretation of the whole being a simple union of the interpretations of the operands. It is only the higher-order operator ('would', in our case) that determines how these alternatives will be exploited in the global semantic import of the proposition, forcing sometimes a disjunctive, but sometimes a conjunctive reading of the complex. These approaches must, however,

[18](Bennett, 2003),p.170
[19](Alonso-Ovalle, 2009), p. 233-236
[20](Lycan, 2001), p.44
[21](Bennett, 2003), p.171
[22](Lewis, 1977), p.361

explain the fact that the disjunction in the antecedent in general seems to allow de Morgan's transformations, as Bennett made clear. And they also have to take good care to prevent the new semantics from validating REa)—otherwise we get stuck again with Antecedent Strengthening.

Once SDA and REa) furnish AS, we can easily show that

(Corner-to-Strict) $A > C \equiv \Box(A \supset C) \equiv A \dashv C$

where $\Box A$ is defined as $\neg A > A$[23], is a theorem of the augmented system.[24] Again, note that the only axioms used in the derivation are fairly trivial—they will be shared by all standard similarity-based counterfactual logics that we shall review here. The upshot is that AS is by and large sufficient to transform > into a strict conditional.

What is more, even in the absence of Replacement of Equivalent Antecedents, SDA will trivialise the modal operators of any strong-centred system (that is, a system validating Conjunctive Sufficiency and Counterfactual Modus Ponens) by making $A \supset \Box A$ logically valid.[25]

4.1.3. Indicative vs subjunctive conditionals

As I have already mentioned, Stalnaker originally intended his theory to provide a viable general framework for both indicative and subjunctive conditionals. Ironically, even though he first grounded his proposal on the intuitive plausibility of the Ramsey test for evaluating indicatives (the selection function was meant to capture on a formal level the speaker's belief revision policies), his theory ended up being considered primarily as a theory of counterfactuals—similarity approaches to indicatives have been found inadequate for a number reasons. We have already seen the trouble in which the so-called Stalnaker's thesis got mired down.[26] In this section we shall have a closer look at why Stalnaker wished to uphold it in the first place, as well as on how he dealt with Lewis's results.

At first, Stalnaker didn't seem to attach much importance to the distinction between indicatives and subjunctives. He considered a possible world to be an adequate 'ontological analogue of a stock of hypothetical beliefs' and was confident that any conflicts between intuitive truth values of conditionals could be accounted for by positing pragmatic shifts in

[23]The truth at α of $\neg A > A$, together with Stalnaker's Constraint 1 guarantees that A should be true across the whole $[\alpha]_R$, which is the standard necessary and sufficient condition for $\Box A$ to be true at α.

[24]See A.1.

[25]The complete proof can be found in (Butcher, 1983), p.75.

[26]See Section 2.3.1.

the selection function.[27] Moreover, at no point did he address Adams's pair and he seems to have regarded the counterfactual form as merely signalling that no credence is awarded to the antecedent of an otherwise standard conditional.[28] Anyway, both indicatives and subjunctives were to behave as ordinary propositions, evaluable at each possible world.

He tried to bolster his semantics by endorsing the Equation[29] as an alternative rendering of the Ramsey Test and by contending that the probabilistic and truth-functional approach could ultimately be made to converge. In Stalnaker's eyes, the fact that both of them yielded the same logic C2 was clear evidence of the viability of his programme. Lewis's triviality results, showing that despite appearances, the two semantic strategies were fundamentally irreconcilable, came therefore as a cold shower.

Unfortunately, even though Stalnaker acknowledged the force of Lewis' proofs and relinquished the Equation, he never seriously undertook the urgent task of reviewing the underpinnings of his approach to indicatives. Given that it initially drew inspiration from the Ramsey Test, but later was found incompatible with a canonical tenet of the latter, an inquiry into what had gone wrong was certainly in order. However, Stalnaker's new take on indicatives left this question completely unanswered.[30]

This time Stalnaker assumed that indicative antecedents forced the selection function to pick out worlds from the closest layer of conversationally live possibilities. This meant that indicative antecedents conventionally signalled their compatibility with the so-called context set comprised by the worlds not ruled out at a prior point in the conversational exchange. Subjunctives, on the other hand, were not supposed to be used unless the speaker intended to reach beyond the shared presuppositions by means of her antecedent.[31]

This admittedly goes some way towards explaining why the inference pattern NAAC

(40) $\neg A \therefore A \rightarrow C$

sounds invalid for indicatives, but since Stalnaker only elaborates on how context sets may shrink throughout a conversation, and never considers

[27] (Stalnaker, 1968), pp. 44-45

[28] Stalnaker (1970), p.120. He did believe, though, that for this very reason, counterfactuals offered us an unperturbed view of the hook-up of an agent's epistemic system.

[29] See Section 2.2.

[30] A passage in (Stalnaker, 1976), p.305 reveals that what Stalnaker most likely had on his mind was some kind of non-epistemic conditional expressing objective causal relations between phenomena. So far as I know, he has never developed this intuition into a full-blown theoretical proposal.

[31] (Stalnaker, 1975), pp.198-202

the possibility of their expanding, there is no explanation for the intuitive invalidity of the related pattern CAC

(38) $C \therefore A \rightarrow C$,

and neither are we presented with any putative explanation of Adams's pair. Unfortunately, Stalnaker's views on the semantics of indicatives grew even more vague and non-committal with the final presentation of his broader theoretical framework in (Stalnaker, 1984).

The seeds of this ultimate failure had probably been sown already in the first article, with Stalnaker's peculiar interpretation of the Ramsey Test. He represented the addition of the antecedent to a belief system, advocated by Ramsey, as a shift from a possible world to another satisfying the antecedent, but otherwise as similar to the original one as it gets. However, as many have noted, a possible world is a far cry from a natural representation of a human belief system.[32] At a possible world, not only all logical and mathematical consequences of a particular proposition are known, but there is no room for any uncertainty about any particular matter of fact. Sets of possible worlds fare much better—even though they are liable to the former idealisation, they do at least eschew the latter. But if we take something like a context set to represent a speaker's (partial) belief system, there doesn't seem to be any distinguished theoretical role for the actual world to play any more—the whole point of the context set is that any of its members could wind up to be the actual world as far as the speaker is concerned.

At this point the theoretical difference between the original Ramsey's idea and Stalnaker's interpretation of it can be brought out. While Ramsey (and following him, Adams, Edgington and Bennett) would say that a speaker will be certain about an indicative $A \rightarrow C$ if all A-worlds from his context set are also C-worlds, Stalnaker would say that this will happen only if revising each and every of the worlds in the context set with A yields a C-world from that very set. In other words, the revision of every possible completion of the speaker's beliefs with A must force another completion supporting C. But this is not how we evaluate the impact of hypothetical new evidence, and thus, given the lesson of the Ramsey Test, cannot be an adequate account of indicative conditionals.[33]

In the rest of the section devoted to similarity, I shall simply assume that Stalnaker's attempt to extend this approach to indicative conditionals is beyond repair and that similarity frameworks are best understood as aimed exclusively at explaining counterfactuals.

[32] See in particular (Gibbard, 1981), p.215.

[33] See also (Lewis, 1976), p.143.

4.2. Generalising Stalnaker

4.2.1. Uniqueness assumption

Perhaps the weakening of $\preceq_\alpha$ that lies most at hand involves allowing for ties between eligible possible worlds, while keeping all other properties of $\preceq_\alpha$ intact.[34] That is, we might want to relax the requirement that there always should be a unique closest possible world supporting the antecedent. This requirement is called the *uniqueness assumption* by D. Lewis.[35] Its abandonment will turn $\preceq_\alpha$ into a *well-founded weak order* with α at its very bottom.

Lewis rejects the uniqueness assumption on the grounds that it requires an excessive degree of sharpness from our notion of similarity. To show that this contradicts our intuitions he makes use of Quine's pair, which we discussed along with Rescher's metalinguistic take on counterfactuals:

(46) If Bizet and Verdi had been compatriots, Verdi would have been French.

(47) If Bizet and Verdi had been compatriots, Bizet would have been Italian.

Lewis's diagnosis is that such examples represent genuine ties—our intuitions about comparative similarity do not favour any of the two conflicting counterfactuals, and they should therefore both be declared false. In our model $\mathcal{M}$, this would mean that the two closest worlds where Bizet and Verdi are compatriots, Italians in the first one and French in the second one, are incomparable as regards their difference from the actual world. If we accept that there is no such unique closest antecedent world, Stalnaker's selection function will remain undefined for the proposition $[\![$'Bizet and Verdi were compatriots'$]\!]$ and we have to come up with another way of specifying truth conditions of counterfactuals.

Fortunately, there is a simple way to achieve this. Informally, the counterfactual $A \boxdot\!\!\rightarrow C$ will be true in α only if there are no accessible antecedent worlds or if the consequent is true in all the closest antecedent worlds.[36] This allows us to simulate Stalnaker's truth conditions by first letting f output sets of closest worlds instead of the single closest one, and then quantifying over this set. So setting our class selection function

[34] Here ties should be construed intuitively—if there is a tie between β and γ, and δ is closer (further) by than any one of them, then δ is also closer (further) by than the other.

[35] (Lewis, 1973b),p.426

[36] For a formal rendering of these truth conditions, see A.2.

$f' : \mathcal{P}(K) \times K \to \mathcal{P}(K)$, $\beta \in f'([\![A]\!], \alpha)$ iff $\beta \in [\![A]\!]_\alpha$ and $(\nexists \gamma \in [\![A]\!])(\gamma \prec_\alpha \beta)$, we obtain the following simpler wording of these truth conditions:

TC Counterfactuals No Uniqueness Assumption II
$[\![A \boxdot\!\!\rightarrow C]\!]^{\mathcal{M}} = \{\alpha \in K \mid (\forall \beta \in f'([\![A]\!], \alpha))(\beta \in [\![C]\!])\}$[37]

Relinquishing the uniqueness assumption also affects the logic. With the new similarity ordering, it is not necessary that C be false across the whole class of the closest A-worlds for $A \boxdot\!\!\rightarrow C$ to fail—a single $A \wedge \neg C$-world among these worlds suffices to falsify it. So it may easily happen that our semantics should not license either of a pair of contradictory counterfactuals, and we must bid farewell to *Conditional Excluded Middle*.[38]

(CEM) $(A \boxdot\!\!\rightarrow C) \vee (A \boxdot\!\!\rightarrow \neg C)$

The very intuitions that were at play in 46 and 47 explain why we might want to get rid of CEM. If Verdi and Bizet had been compatriots, would Bizet be Italian or not? Neither answer sounds convincing and we are tempted to reject the question altogether, as there doesn't seem to be any way to decide one way or the other.

Stalnaker doesn't question these intuitions, but disagrees with Lewis about how they should be incorporated in a theoretical account of counterfactuals.[39] For Stalnaker, no semantic theory can avoid certain theoretical idealisations that have to be relaxed in practice. His well-ordering of possible worlds purports to be exactly such an idealised ingredient, whose neatness often sets too high a bar to clear for practical purposes. He argues that rather than by implementing our intuitions directly into the semantic wiring of counterfactuals, they are better explained by resorting to van Fraassen's supervaluation theory of vagueness. Supervaluations allow us to evaluate linguistic items in spite of their underdeterminacy in a context, by looking at the class of all possible resolutions of such underdeterminacy. A long story short, the upshot is that (46) and (47) come out as neither true nor false on this account, since different, for all we know equally plausible ways of completing the similarity well-ordering give rise to different truth values. This, admittedly, sounds more palatable than to declare them outright false. After all, we seem to accept

[37] The vacuous case is taken care of by stipulating that $f(\emptyset, \alpha) = \emptyset$ for all α.

[38] The parallel between CEM and the uniqueness assumption is not perfect, though. For CEM to be valid, the underlying ordering must be a well-order, so if the limit assumption fails, CEM will fail irrespective of the uniqueness assumption.

[39] (Stalnaker, 1981),p.89

(56) If Bizet and Verdi had been compatriots, then Verdi would have been French or Bizet would have been Italian.

Lewis admits that (56), together with the negations of (46) and (47), sounds like a genuine contradiction.[40] He also acknowledges that the resort to supervaluations by and large solves the problem of implausibly high demands upon our capacity of determining the similarity ordering. However, he still objects to Stalnaker's theory on the grounds that it cannot account for *might-counterfactuals*.

4.2.2. Interlude: Might-counterfactuals

The rejection of the uniqueness assumption presented Lewis with an opportunity to define the might-counterfactual as a dual operator to the would-counterfactual. Since their relation, as construed by Lewis, is highly reminiscent of that of simple possibility and necessity, Lewis represented them respectively with a diamond and a box attached to an arrow: 'If A had been the case, then C might have been the case' became $A \diamond\!\!\rightarrow C$. The might-counterfactual is defined as follows:

(M-W) $(A \diamond\!\!\rightarrow C) \equiv_{def} \neg(A \mathrel{\square\!\!\rightarrow} \neg C)$

This has powerful backing in our intuitions: we want to reject 'If Verdi and Bizet had been compatriots, Verdi would have been French' precisely because in that event, Verdi might not have been French, but rather Italian, just as Bizet. The corresponding truth conditions (in terms of the class selection function f') are as follows:

TC Might-Counterfactuals No Uniqueness Assumption
$[\![A \diamond\!\!\rightarrow C]\!]^{\mathcal{M}} = \{\alpha \in K \mid (\exists \beta \in f'([\![A]\!], \alpha))(\beta \in [\![C]\!])\}$

As we have seen, Stalnaker advocates the uniqueness assumption and CEM, which under the current definition would equate might- to would-conditionals. Lewis retorts that we do not seem to have any other plausible treatment of might-counterfactuals at our disposal that could do justice to our pre-theoretical intuitions, if 'might' is to be construed in terms of the standard objective possibility operator $\Diamond$:

> Four candidates come to mind: $\Diamond(\phi \& \psi)$, $\Diamond(\phi \mathrel{\square\!\!\rightarrow} \psi)$, $\phi \mathrel{\square\!\!\rightarrow} \Diamond\psi$, and $\phi \mathrel{\square\!\!\rightarrow} \Diamond(\phi \& \psi)$. But none will do. Take ϕ as *'I looked in my pocket'* and ψ as *'I found a penny'*; suppose I did not look, suppose there was

[40] (Lewis, 1973a), p.80

> no penny to be found and make commonplace assumptions about relevant matters of fact. Then *'If I had looked, I might have found a penny'* is false, but all four candidate symbolizations are true.
>
> (Lewis, 1973a), p.80

This presses an important point about our use of modalities. From a technical point of view, Lewis is of course right, but only to the extent that $\Diamond$ is a highly abstract theoretical device, which only imperfectly reflects our ordinary use of 'may' and 'might'. We don't expect the latter to quantify over the whole $[\alpha]_R$ even in their alleged 'objective' use.[41] Typically, when I say 'You might have got sick!', I am not trying to convince you that it was metaphysically possible in the broad sense—you already know that much (even though, having opted for a more useful occupation in your life than me, you may never have actually heard of metaphysical possibility). What I am getting at is that it was possible *in the circumstances*, and that is exactly how 'might' works also in counterfactual contexts.

Stalnaker correctly notes that 'might' plays the same semantic role in conditional and non-conditional natural language contexts, but concludes that it is because this role is epistemic rather than ontic, and that its scope is the whole counterfactual, rather than just the consequent.[42] Neither claim is easy to put to the test. The first one, because we have no clear idea of how uncertainty about a counterfactual would differ from conditionally acknowledging a certain situationally bounded ontic possibility, as adumbrated above. And the second one, because we have no clear idea of how uncertainty about a counterfactual would differ from uncertainty about a proposition under a counterfactual assumption. I will argue that these distinctions are spurious, but in any case, the natural language 'might' cannot simply be the $\Diamond$ of our standard modal logic.

Stalnaker's diagnosis is that there are two clashing readings of the modal involved in Lewis's example: an epistemic one, under which the counterfactual 'If I had looked, I might have found a penny' is true, and a quasi-epistemic one (expressing compatibility with speaker's knowledge enriched with all the relevant facts), under which it is false. I would like to suggest that the first reading is a theoretical fiction which doesn't capture how modals are really used in natural language. When I utter 'If ϕ, then might ψ', I am not simply making a report about my belief state—if that were so, you couldn't correct me by pointing out that there were no pennies in my pocket (something I didn't know). The second reading is intuitively

[41] I shall argue in Chapter 7 that the sharp difference between epistemic and ontic modalities is a by-product of theorisation, without a clear counterpart in natural language.

[42] (Stalnaker, 1981),p.99

more acceptable, but leaves open the question of how the set of relevant facts should be determined. And then, could anyone ever be sure that she knows all the relevant facts, whatever they might be? If not, then no assertion of such a modal sentence will be warranted, making it rather useless for a natural language user. It is clear that an altogether different approach is needed to account for the interaction between modals and conditionals.[43]

4.2.3. Strong centring

We still have the actual world α standing alone at the very bottom of our relation $\preceq_\alpha$, a property called *strong centring* in the literature. While admittedly no other world can resemble α more than α itself does, one can argue that there may be worlds which differ from α only in such irrelevant respects that they should be considered for all intents and purposes indistinguishable from it. This would amount to relaxing the strong centring to a mere *weak* one. As far as I know, no one has ever argued against weak centring.

This further generalisation, however, doesn't require any change in our new truth conditions, as it only amounts to allowing a tie also at the bottom of the ordering. It does affect the logic, though—*Conjunctive Sufficiency* will no longer be valid.

(CS) $\quad (A \wedge C) \supset (A \mathrel{\Box\!\!\rightarrow} C)$

As you may already have guessed, it is no particularly dramatic loss. CS sounds awkward, because almost all counterfactuals with a true antecedent do. However, and for the same reason, it doesn't sound completely wrong either. Our intuitions simply provide no reliable guide in such matters—that's why we may just as well let CS go. With weak centring, we shall still retain *Conditional Modus Ponens*.

4.2.4. Limit assumption

Lewis also rejected the principle for which he coined the name *limit assumption*. In a sense, it is a complementary constraint to the uniqueness assumption—whereas the latter ensures that there will always be at most one closest antecedent world, the former secures the existence of *at least* one. Antecedents that apparently fail to deliver any closest antecedent worlds are exemplified by counterfactuals such as

[43] See Section 5.2 and Chapter 7.

(57) If I were taller than I am, I could touch the ceiling.

Is there any minimal increment in my height which one should consider when entertaining the antecedent? Arguably not. For any hypothetical increase d in my height, there is a smaller one, say $d/2$, which will make my counterfactual height closer to the actual one, and thus preferable from the point of view of the present theory. The upshot is that there is no set of closest antecedent worlds. Abandoning the limit assumption opens the possibility of $\preceq_\alpha$ not being well-founded, or equivalently (with the Axiom of Choice), the possibility of infinite chains of possible worlds coming ever closer to the actual world.

It is relatively easy and even rewarding to shed uniqueness, as it makes the system enormously more flexible and interesting, while keeping it quite simple at the same time; however, allegiance to the limit assumption among scholars has proved rather resilient. One of the reasons is that the generalisation of the elegant truth conditions we have been dealing with up to now becomes somewhat unwieldy in the new system.

Informally, the counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ will be true in a possible world only if there is no accessible A-world or if there is an accessible A-world below which all accessible A-worlds are also C-worlds. And the might-counterfactual $A \mathrel{\Diamond\!\!\rightarrow} C$ will be true in a possible world only if there are accessible A-worlds and each of them is either also a C-world, or has such a world below itself.[44]

This formulation of truth conditions makes it impossible to define any straightforward selection function on Stalnaker's lines, not even a class one, simply because in general nothing guarantees the existence of the set of closest antecedent worlds.[45]

But how plausible is the abandonment of the limit assumption anyway? Let us return to (57). According to Lewis, our intuitions cannot supply any set of closest antecedent worlds, so in a system with the limit assumption we will either be compelled to choose one *ad hoc*, or regard it as the empty set, in which case the counterfactual will be trivially true. Lewis strives to avoid this outcome by giving up the assumption and adjusting the semantics accordingly. With these changes, however, if in the actual world (measuring, say, h cm) I am too short to reach the ceiling, then (57) comes out strictly false. That's because if $h + d$ cm is the minimal height required for reaching the ceiling (let us suppose that the length of my arms is a function of my overall counterfactual height), then the world in which I

[44] For a formal rendering of these truth conditions, see A.2.

[45] See (Lewis, 1973a), pp. 57-60.

am $h + d/2$ cm tall will be closer than the world in which I am $h + d$ cm. It might seem that we have jumped out of the frying pan right into the fire.

However, bad though this may seem, there is always the Gricean cavalry to save the day. It can be argued that since examples such as (57) all convey literal falsehoods, which is common knowledge between the speaker and her audience, the latter is invited to pragmatically firm up the antecedent with a non-articulated item, such as:

(58) If I were *a little* taller than I am, I could touch the ceiling.

Now, 'a little' is still fairly context-dependent for its interpretation, but from the point of view of the limit assumption, it is a game-changer. Whatever may be considered as 'a little' under the circumstances, there is no longer an infinite descending chain of antecedent worlds ever more closely resembling reality.[46] We are back in the calmer waters where a set of closest possible worlds can be identified.

This seems to be a general pattern among counterfactual examples put forward to illustrate the need for jettisoning the limit assumption—arguably, if the antecedent worlds get closer and closer to the actual world, at certain point it will be no longer in our power to appreciate the difference between a counterfactual scenario which still supports the antecedent and one that already ceases to do so. While trivial counterfactuals such as

(59) If I were taller than I am, I would measure more than h cm.

(where h, as before, is my actual height) will still count as true, any similar counterfactual exploiting a non-trivial consequence of my height will in all likelihood turn out to be unacceptable on this semantics. This is further illustrated by the fact that

(60) If I were taller than I am, I would measure less than $h + d$ cm.

will be true for all $d > 0$, so d will approach 0 in the limit. Now, if you combine the consequent of (59) with all the instances of the consequent of (60), it is intuitively clear that you will obtain an unsatisfiable set of sentences.[47] So if you are uncomfortable with the idea that a consistent proposition may yield an inconsistent set of counterfactual consequences, you will certainly want to get the limit assumption on board.

In view of these examples, it is to be expected that the antecedents of most counterfactuals for which the presence of infinite descending chains

[46] By the way, it may be worth pointing out that it usually cannot simply be interpreted as 'd, whatever it might be', as not all the likes of (57) strike us as trivially true.

[47] See (Herzberger, 1979) for details of the proof.

of antecedent worlds makes any difference will eventually have to be strengthened by the audience to a form where a set of closest antecedent worlds can be found. The literal import of such counterfactuals will correspond only to a intermediate stage of their processing, triggering an implicature reflecting the real intended meaning of speaker's utterance.

Pragmatic enrichment is a tricky phenomenon, which can occur for a number of reasons. It may prove difficult to chalk it up to one particular aspect of an utterance. For instance, one may argue that some kind of pragmatic reinforcement arises by default in many typical comparative assertions: if I say that Juan is taller than María in an informal setting, I am usually expected to disregard height differences below a certain level. You may therefore be able to argue that the requisite enrichment takes place already before the semantic mechanism of counterfactuals kicks in.[48] Alternatively, you can admit that the enrichment sometimes *is* triggered by the absence of an eligible set of closest antecedent worlds, but declare yourself interested in modelling exclusively the final result of this process. Either way, you will have a strong case for retaining the limit assumption.

Another reason for maintaining the limit assumption is that its logical cost is very subtle, if what you are interested in is the study of counterfactual reasoning in natural language. The limit assumption is necessary for the validity of CEM, but once the uniqueness assumption is relinquished, it doesn't go hand in hand with any other axiom of C2—its addition to the system only makes a difference for certain arguments with infinite premises. To stick with it, you will have to sacrifice compactness.[49] While this severely compromises the completeness of your logic, as you never really use such arguments in real life (do we ever use more than two or three premises in our daily counterfactual reasoning?), you may well find the bargain worth your money.

4.2.5. Almost-connectedness

Let us take stock. Our similarity ordering $\preceq_\alpha$ is now a mere partial order, with α as one of the minimal elements, and there can be ties between possible worlds for closeness to α. But how did we construe the ties? Any

[48] This is roughly the view taken in (Warmbrōd, 1982). Warmbrōd also argues that in the special case when the speaker forces an interpretation at odds with the limit assumption, he will end up trivialising his counterfactual, as the antecedent will turn out to be unsatisfiable. Stalnaker, for his part, contends that the use of counterfactuals with such antecedents is inappropriate and the selection function will be undefined for them: (Stalnaker, 1981), p.97.

[49] (Veltman, 2012), p.4

worlds incomparable by means of $\preceq_\alpha$ were considered to be equally close to the actual one, so if any world was closer (further away) than any of those incomparable worlds, it was closer (further away) than all of them. This allows us to partition the universe in equivalence classes of equally close worlds and regard (by means of an induced ordering, of course, but let us be a little sloppy for simplicity's sake) $\preceq_\alpha$ as a linear order on this partition.[50] Accepting the limit assumption turns this linear order into a well-order.

Should we want to reduce $\preceq_\alpha$ to a mere partial order, without any additional properties, we would have to give up both weak centring and the possibility of performing the aforementioned partition.[51] The latter modification amounts to taking back the requirement that the relation of being 'equally' close to actuality should be an equivalence relation. As shedding reflexivity or symmetry would stop $\preceq_\alpha$ from being a partial order, it is transitivity that must be relinquished. One can interpret this move, following (Lewis, 1981), as an admission of genuine incomparabilities alongside mere ties.[52]

Requiring $\preceq_\alpha$ only to be a partial order means sticking exclusively with the axioms CI,CC, CW, ASC, AD and RE, together with all propositional tautologies and the rule of Modus Ponens; that is, with the table on page 75 up to AD. This reduced system gives us the simplest logic for similarity counterfactuals, sometimes called P after Pollock.[53]

4.3. Possible worlds

After mastering the two preceding sections, the reader probably can guess why Stalnaker-Lewis's systems immediately proved so successful among logicians and formal linguists—the formal framework is mathematically beautiful, but perspicuous enough for a layman to follow. Earlier theories, even if they sometimes took a stance on a point of logic, were in no condition to provide anything similar.

However, time has come to subject the theory to detailed scrutiny. You have been promised testability on the basis of the logic brought forth by the semantics, but the purpose of the last section was to bring home to you the extent to which this logic is malleable—little amendments to the

[50] These equivalence classes correspond exactly to the layers of Lewis's spheres in his original formulation of the theory in (Lewis, 1973a).

[51] This is the system favoured in (Pollock, 1976).

[52] See A.2 for details on how this will affect truth conditions and logic.

[53] (Veltman, 2012), p.3.

system can give you just the axioms you ask for. Reasoning in natural language is anything but systematic, and any logical system that purports to capture its structure will be compelled to resort to idealisations. Often there is simply no conclusive linguistic evidence in favour of or against implementing a rule of inference. The distinction between semantic and pragmatic phenomena is always drawn against a fairly theoretical background.

We have, therefore, to look at the underpinnings of the similarity theory in order to see how it fares not only on the structure, but also on the content of counterfactual reasoning. There are two basic elements to the evaluation of particular counterfactuals—the universe of possible worlds and the relation of resemblance between them. We'll start with the first one and discuss several different construals of their role in formal semantics.

I shall argue that despite appearances to the contrary, Lewis's realism has cast a long shadow. Most scholars were so eager to reject it that they didn't realise it was put forth as an answer to a genuine question, to wit: *What is the purport of modal talk?* Having marshalled a few arguments against realism and conceiving possible worlds as a convenient fiction rather than 'real chunks of rock', they went on doing basically the same kind of semantics as Lewis, mostly unaware of the fact that they still owed us an important explanation. Even worse, having rejected Lewis's eccentric metaphysics, they tacitly embraced the erroneous picture of language which had given rise to it. The upshot is that despite all the sophistication of present-day possible world semantics, its achievement mainly resides in providing a perspicuous framework for *rephrasing* modal talk, rather than in helping us *understand* the role modal expressions play in natural language. In my view, here lies one of the sources of the scholasticism and intellectual constipation that hold sway in large areas of modern linguistics and philosophy of language.

4.3.1. Realism

Realism about possible worlds is just what it sounds like—the thesis that when we use modal vocabulary, we are talking about huge concrete objects, among which our actual world takes no pride of place. Modal expressions, on this account, work similarly to the adverbs of place: their absence suggests that the speaker is talking about the place of utterance, while their presence switches the circumstance of evaluation to the location thus indicated (albeit often only through a quantificational phrase). Analogously, non-modal utterances are interpreted as being about the actual world,

while the addition of a modal operator results in the same affirmation being made of some possible world(s) that may not coincide with the actual one.

Philosophers are usually not the kind of folk that shies away from wacky theories, but realism is apparently such a bizarre position that it has become firmly identified with D. Lewis, its original and almost sole advocate.[54] Were it not for Lewis's philosophical eminence as well as his witty and engaging writing, realism might well have remained at the fringes of the philosophical debate. In what follows, I shall not take great pains to argue exhaustively against realism. I submit that if we are ever justified in cutting any corners, it is here; I shall by and large take the untenability of realism for granted. Realism itself is not a difficult target (I surmise Lewis may have been prodded into its adoption by the very challenge of defending it)—it is rather the common alternative practice of treating possible worlds as an unexplained explainer that needs to be analysed with care. I will be more interested in why realism emerges in the first place and I'll try to find out whether alternatives to realism can allay these original preoccupations.

Let us briefly review what pretty much everyone seems to take to be the decisive argument against modal realism—the epistemological objection. It goes roughly as follows: we get to know real objects by entering into causal relations with them. But this is by definition ruled out in the case of possible worlds—it is their very causal disconnection from the actual world which makes them merely 'possible'. So how could we ever learn

[54]The only other realist reported in (Bennett, 2003) is R.B. Miller. He advocates what in (Miller, 2001) is called 'moderate modal realism', but the 'moderateness' only concerns the range of entities inhabiting possible worlds, not the actual ontological status of the latter. Another scholar defending a position close to Lewis's realism is J. Divers. In his (Divers, 2002), he distinguishes 'genuine realism' from 'actualist realism', which endorses the existence of possible worlds, but identifies them with some other (possibly abstract) actually existing entities. After evaluating both against a number of criteria, he concludes that preference should be given to the former (without prejudice to a possible 'antirealist' victory, out of the monograph's compass). I find Divers's evaluation criteria too technical and little related to the task of explaining modal discourse. What is more, I don't see the dilemma in either Miller's or Divers's terms—rejecting possible worlds as existing concrete objects, while granting them a seal of approval as abstract entities, doesn't require identifying them with some other, *prima facie* more respectable objects. A. Kratzer vows allegiance to Lewis's construal of possible worlds in a few terse remarks (see for instance (Kratzer, 1977), p.10), but to my knowledge never really presses the point, leaving foundational issues for all intents and purposes outside the scope of her theorising. For F. Veltman, possible worlds are just unanalysed set-theoretical objects that represent or model situations conceived of by a speaker: (Veltman, 1976), p.251.

anything about them?[55]

Lewis replies that the epistemological premise must be rejected anyway, not only for the sake of retaining his modal ontology, but also in order to account for mathematical knowledge. We do not have causal transactions with numbers or sets, but we do accept their existence, on pain of radically changing the subject matter of mathematics.[56]

This reply, in my view, falls prey to a singular confusion. Lewis strives to brush off the difference between abstract and concrete objects, which others rely on in order to explain the peculiar behaviour of numbers. But the distinction is intuitively clear enough—a cat or a hammer can, just as the empty set or a schedule cannot, startle you by unexpectedly falling on your head and cause you severe injuries.[57] As biological creatures, we have to find our way in a physical world full of concrete objects, but so as to keep track of their properties, we can enrich our conceptual domain with abstract ones at our discretion, as long as they serve the purpose for which they were introduced.[58] The case of the foundations of mathematics is especially instructive: how much choice should we be allowed to use? Does Ω of the non-well-founded set theory, as opposed to, say, the more familiar $\emptyset$, *really* exist?[59] And is there any principled way of answering such questions?

The so-called ontological questions have, of course, bedevilled philosophy from its very beginnings and platonism survives in one form or other to this day. However, platonism cannot simply amount to the thesis that abstract objects exist—if this were so, everybody would become a platonist immediately upon momentarily abandoning the conceptual rigours of their philosophical system in order to attend to more mundane issues. In addition to believing in the existence of abstract objects, a real platonist must construe them along the lines of concrete objects—as if they were

[55] In the same vein, D. Sanford convincingly argues that even if concrete objects analogous to possible worlds really existed, they could not have any bearing on our language practice: (Sanford, 1989), pp.156-172.

[56] (Lewis, 1986a),p.109

[57] G. Ryle made this point more brilliantly than I could ever do throughout his (Ryle, 1949). The fact that a mere quarter of century later platonism was again in the centre of philosophical debate reflects poorly on analytic philosophy's capacity to learn from the great thinkers of its own past.

[58] This invites the following empirical speculation: could it be that the more advanced cognitive processes associated with abstract thinking simply made use of the pre-existing module used for handling concrete objects?

[59] Remember that in the non-well-founded set theory, Ω is the set whose only element is itself. Also, before you dismiss such set theory as nonsense, note that it has important computational applications.

hovering in a distinctive realm (as 'raisins in a pudding'[60]) accessible only to human reason through some kind of intellectual apprehension, modelled upon ordinary empirical inspection. A real platonist, such as Gödel, believes that there is *only one correct answer* to the continuum hypothesis, because if only you could clearly inspect the objects corresponding to $\aleph_1$ and $\beth_1$, you would know what it is.

This is, by the way, also why in my view attributions of platonism to Frege either seriously miss the point or border on triviality. Frege, of course, accepts the existence of abstract objects and is very explicit about how we can use them to our advantage in the construction of a mathematical and semantic theory, but as far as I have seen, he never indulges in the kind of intellectual daydreaming characteristic of real platonists.

Platonism is mostly innocuous as far as mathematical practice is concerned, and in some cases it may even provide a boost to a mathematician's imagination. It fails, though, as an *account* of such practice. Mathematical, and in general abstract objects are to a large degree constructed, not discovered, by us. Of course, there are constraints to what we reasonably can postulate, but they are imposed by the purpose for which the abstraction is deployed in the first place. If your arithmetic gives you '3=4' as a theorem, you won't be able to use it in a marketplace, as four apples are more valuable than three. How much choice you choose to use may have interesting consequences in advanced set theory, but as long as it lets the calculus go through, engineers will not bother to object.

The same holds for possible worlds—Lewis's realism doesn't preclude him from using and analysing modal vocabulary just as his fellow semanticists do and may even be one of the secrets behind the virtuosity of his analyses. However, it falls short as an attempt to explain what we are really up to when using modal operators in natural language. To borrow a felicitous phrase by Brandom:

> The trouble with taking it that there is something that is successfully represented by each purported representing is not just that it involves commitment to a luxuriant ontology; ontological self-indulgence is a comparatively harmless vice. But it can be symptomatic of a failure to shoulder an explanatory burden.
>
> (Brandom, 1994), p.71

Brandom addresses a different issue here, but the point is completely general (one can even argue that for a platonist, possible worlds play precisely the role of designata of hedged representings). Accepting all

[60]The fortunate simile is due to Larry Powers, quoted in (Stalnaker, 1984), p.43.

kinds of new suspicious concrete entities is a common strategy in order to bypass a request for an explanation of the import of a problematic piece of discourse. A flamboyant ontology is often a price for keeping the underlying theory of meaning simple—after expanding the domain of existent concrete objects, any such discourse can be simply brushed aside as unproblematically *descriptive*.

However, the fact that such an explanation is unavailable or at least highly unsatisfactory in the case of abstract objects is just one of their hallmarks, and by now we know why that should be so—their properties are to a large extent determined by us and not by the world. That's why I find the price for an unfettered descriptivism to be too high, and especially so in the case of possible-world realism—it prevents us from making sense of modal discourse within a picture of language as a biological feature of the human animal. There is no place for possible worlds as concrete objects in such a picture and thus no use for the language game of describing them.

4.3.2. Possible worlds in semantic analysis

Once we have put possible worlds in their place, dissolving any remaining doubts as to their ontological status as abstract objects, the question arises about the language game they hail from. What I have in mind is the origin of the identity criteria which enable the introduction of abstract objects as objects in the first place. In familiar cases of abstract objects we can at least gesture towards the language practice that is fostered by outsourcing one of its elements to the abstract realm. The example of mathematics will, once again, prove instructive.

In unstudied natural language numbers function exclusively as determiners: there are no 'ones' or 'twos' by themselves, but only 'one goat' or 'two microwave ovens'. And this is just as well when there is not much more around than a mere bunch of things of any given kind. However, when trade carts eventually get trundling between villages and, much to our chagrin, taxes become collected, processing figures starts exacting too high a toll on our mental resources. The solution is to separate them from anything that doesn't play any substantial role in the requisite calculations—and this is how mathematics and with it, the number as an abstract object, is born.[61]

[61] However, when platonist tendencies succeed in making the abstraction self-serving by severing all its ties with the original practice, the abstraction may eventually become unsustainable and implode. I suspect this is one of the reasons behind the spectacular failure of the 'New Math' programme.

As with numbers, so with voices, predilections, institutions, intentions and itineraries. In each case we can readily think of a down-to-earth rational activity involving exclusively concrete objects, upon which the abstract talk indirectly supervenes. Note that the direction of explanation is invariably from the abstract talk to the concrete one. Thus a parent faced with the question 'What is "bias", mum?' will probably answer something like 'That's when you treat someone unfairly, darling.' Observe as well that such an explanation, while being perfectly adequate on its own merits, doesn't provide a reduction of the sort that would satisfy a Vienna Circle member. For a number of reasons, I think we should probably withstand the temptation of a general project of reducing abstract objects to some better-behaved entities.

First, I am perfectly happy with abstract objects as they stand (and so are most people, I should dare to suggest, as long as the word 'ontology' doesn't make its entrance. Apparently, it is capable of scaring a person into denial about her most innocuous habits.). Second, the spectacular failure of reductionist programmes in the 20th century philosophy should make us wary of such an undertaking's chances of success. I think we had better acknowledge the lesson—the posteriority of one language game to another doesn't necessarily imply mutual interdefinability of their vocabularies.

Despite these *caveats*, there is no denying that sometimes a convenient paraphrase is readily available—the extreme case being theoretical notions constructed by explicit definitions.[62] This is also the case with Lewis and Stalnaker, both of whom introduce a possible world as 'a way the world might have been'.[63] The difference between them boils down to the fact that while Lewis conceived of these ways as concrete giant objects, Stalnaker explicitly rejected any attempts to recast this turn of words in more tangible terms. For him, as for us, possible worlds are 'abstract objects, whose existence is inferred or abstracted from the activities of rational agents'[64]. Only when the role that entertaining such possibilities plays in rational agency has been specified will the analysis of modal notions in terms of possible worlds really succeed in its explanatory task. In the meantime, our comprehension of possible worlds will piggyback on our previous modal intuitions, conjured up by the phrase 'the way the world/things might have been'. Something similar must have been on van Fraassen's mind when he characterised his stance on possible worlds semantics:

[62] For instance, the notion of a cardinal number sometimes is defined before introducing ordinals by stipulating that card A = card B iff A and B are equipollent.

[63] See (Lewis, 1986a), p.2 and (Stalnaker, 1984), p.45.

[64] (Stalnaker, 1984),p.50

> I see the possible worlds machinery just as Duhem saw the rope-and-pulley models of the English physicists: such fictions are useful when giving an account of the surface phenomena—and there is, in reality, nothing below the surface. In our case the phenomena are the inferential relations among statements, attested in the inferential behaviour of those engaged in such discourse.
>
> (van Fraassen, 1976), pp.266-7

It is worth pointing out that an explanation of possible worlds (and thus of our rudimentary modal operators) based on rational human activities exceeds the scope of traditional semantic analysis. Rather than a specification of truth conditions, it supplies a description of a practice which makes clear why and how modal vocabulary is actually developed.

Both Lewis's extreme realism and the shoulder-shrugging attitude typical of contemporary formal semantics plead for an exemption from the obligation of providing an account of possibilities along Stalnaker's lines. And there is strictly speaking nothing wrong with such an approach, as long as their proponents acknowledge its essential incompleteness. However, such candidness is not very common in the literature. Lewis, of course, regarded modal talk simply as descriptive, so he felt no urge to supplement his account; but most semanticists, without having recourse to Lewis's excuse, follow him in their unworried treatment of possible worlds as the fundamental building blocks of their theories. Hence the most they can aspire to is to define some more complex modal vocabulary in terms of bare possibility. Of course, such analyses are not always without merit (provided that their other semantic ingredients are non-trivial). On the downside, however, they keep us in the dark about the ultimate purport of modal talk, while coming to grips with it might well supply the understanding necessary for linking the analysed expression with the analysandum. This is how it can happen that after successfully slogging through many a virtuosic semantic analysis, we have no clue as to what function the analysed expression might possibly play in our linguistic lives.

Furthermore, it is usually assumed that modal logic models are theoretically neutral with respect to the study of modalities, perhaps because of the very triviality of the definition of a possible world. However, in Chapter 7 I shall argue that this might not be so. Just as any other scientific model, Kripke frames are constructed upon theoretic pillars which to a certain degree pre-shape the analyses carried out within the framework. This is most notoriously reflected by the problems besetting their treatment of mathematical knowledge as well as the implementation of quantified modal logic. To pin down the limitations of this apparatus in modelling

natural language, I think our attention must go beyond philosophical uses of modal statements, which are technical and comparatively rare and so can easily lead us to error about what we normally use such statements for.

4.4. Similarity

Time has come to shed some light on what can be properly regarded as the linchpin of Stalnaker-Lewis analysis—the notion of similarity underlying our ordering $\preceq_\alpha$. Up to now, we have only examined its structural properties and showed how tinkering with some of them translated into modifications in our conditional logic. These structural properties, of course, only fix the truth values of a very reduced number of mostly trivial counterfactuals, such as $A \boxrightarrow A$ or $(A \wedge B) \boxrightarrow B$. To evaluate any really interesting counterfactuals, we shall have to flesh out the details of our ordering.

The relevant notion of similarity is meant to guide and inform this process. It is not necessary for us to have clearer intuitions about the requisite kind of similarity than about actual truth values of counterfactuals—it may be one of those notions that are best explained by putting them to work. Nonetheless, I do think that it should be possible to get some independent grip on this concept if the theory purports to provide an *explanation* of our use of counterfactuals. The notion of similarity being central in determining truth values of counterfactuals, a coherent picture of the former should enable us to understand what we are up to when entertaining and asserting the latter.

For all what has been said, there is a wide consensus in the literature that Stalnaker-Lewis theory, as it stands, fails to supply an adequate criterion of similarity, which leaves it toothless when it comes to explanatory or predictive tasks. In this section I shall explain how this came about and what the consequences are for our search for theoretical alternatives.

4.4.1. A pragmatic question?

We have seen that Stalnaker introduced the notion of minimal difference from the actual world in order to simulate on the ontological level the minimal belief revision mandated by the Ramsey test for evaluating indicative conditionals.[65] However, the guidelines for this minimal revision turn out

[65](Stalnaker, 1968),p.45

to be at least as elusive as those for the pertinent similarity comparisons. Seldom are we told more than a few truism about 'accommodating the antecedent' or not making 'gratuitous changes to our belief system'. For the account to really serve as an explanation of natural language conditionals (Stalnaker's self-confessed goal), the notion of 'resemblance' must be made clearer.

Instead of doing that, though, Stalnaker resorted to a classical move in language analysis handwaving—he banished the phenomenon to the realm of pragmatics. The specification of the selection function f for conditionals, the argument goes, is akin to the domain restriction for quantified sentences. It varies from one context to another and is usually tacitly shared among the participants in a conversation. According to Stalnaker, the actual value of f will depend upon (presumably among other things) 'the context of utterance, the purpose of the assertion and the beliefs of the speaker or his community'[66].

This is redolent of the strategy pursued by the liberal wing of cotenability theoreticians[67] and it won't do for exactly the same reasons. Our room for manoeuvre when interpreting conditionals is not as wide as the pragmatic picture seems to suggest. I am afraid that the analogy with quantifier domain restriction cannot withstand closer scrutiny.

First of all, in the event of a communication failure due to diverging domain assignments by participants in a linguistic exchange, there is always the option of making the domain explicit. Yet it is by no means clear how such a clarification would look like in the corresponding case of a conditional.

Second, settling on the domain only fixes the truth conditions of a quantified sentence, not its truth value. It still remains to be checked whether the truth conditions really obtain. However, specifying f is enough for evaluating a conditional. To preserve the analogy, the context should not contribute the entire f, but only some rule for ascertaining it.

And third, even in the absence of the above problems, the main point made about the metalinguistic liberal wing applies also here—placing a language phenomenon on the pragmatic side of the divide doesn't exempt us from the duty to come up with an account of the phenomenon. Admittedly, to say that the quantification domain is in each case restricted to the set of contextually salient or otherwise relevant objects is not to say very much, but we can still make much better sense of what it means to be contextually relevant for an object than for a similarity ordering of possible

[66](Stalnaker, 1968), p.51

[67]See page 51.

worlds. I suspect that the latter notion is incomprehensible without further elucidation.

4.4.2. Lewis's take on similarity

From his remarks in (Lewis, 1973a) it clearly transpires that, as opposed to Stalnaker, Lewis had something very close to the everyday notion of similarity on his mind when he first advanced his theory of counterfactuals:

> [...] any problems posed by my use of comparative similarity differ only in degree, not in kind, from problems about similarity that we would be stuck with no matter what we did about counterfactuals. Somehow, we *do* have a familiar notion of comparative overall similarity, even of comparative similarity of big, complicated, variegated things like whole people, whole cities, or even—I think—whole possible worlds.
>
> (Lewis, 1973a), p.92

Even though Stalnaker wasn't able to spell out the notion of similarity he relied on in any detail, he acknowledged its technical character by linking it with the 'methodological policies' for modifying our beliefs in the presence of new evidence, required for the Ramsey test. Lewis, on the other hand, took 'resemblance' quite literally—as he conceived possible worlds along the lines of concrete objects, the option of making use of our unstudied intuitions about similarity between such objects was open for him.

More than the question as to whether our everyday notion of similarity was germane to counterfactual reasoning at all, it was its vagueness that most occupied Lewis. Against those who might be inclined to reject his account on the grounds of the high context-dependency of the concept of similarity, Lewis argued that given the widely acknowledged vagueness of counterfactuals, it was only desirable to analyse them in terms of another vague notion and make them 'sway together'[68]. Take another example of Quine's, this time about the Korean war:

(61) If Caesar had been in command, he would have used the atom bomb.

(62) If Caesar had been in command, he would have used catapults.

[68](Lewis, 1973a), p.92

That each of this pair of clashing counterfactuals can separately be deemed reasonable is, according to Lewis, due to a switch in the vagueness resolution of our notion of similarity[69]—we are inclined to accept (61) focusing exclusively on Caesar's temperament, while with (62) the relevant comparison criteria are extended to cover also other traits of Caesar's personality, such as the actual limits on his knowledge of the art of war. We are not obliged to stick with any one resolution of such vagueness, of course, as long as we measure up to the reasonable expectations of our interlocutors.

Nevertheless, most attacks came from a thoroughly different angle. Some of them complained that Lewis's theory couldn't account for counterfactuals stating that a tiny alteration of a certain condition would have brought about a significant change in the unfolding of the events. They typically follow the pattern of Kit Fine's nuclear holocaust argument:

(63) If Nixon had pressed the button, there would have been a nuclear holocaust.[70]

Suppose there is, fortunately, never going to be a nuclear holocaust as things stand. Then Lewis's theory predicts (63) to be false, despite our intuitions to the contrary, because a world β where, say, the wiring of the button fails and the Earth as we know it is preserved, will on any account be considered more similar to the actual one than any other world γ where the holocaust does occur (here it is essential that the incident that restores the normal course of events at β be of marginal importance as regards similarity comparisons with γ). But sometimes we definitely do want to say that certain events would have seriously derailed the progression of the world history.

A related criticism also concerned the weight of actual matters of fact in similarity assessments. This time, however, the inclusion of the offending fact in the counterfactual scenario was deemed gratuitous rather than outright illicit. The example is by P. Tichý:

> [...] consider a man —call him Jones— who is possessed of the following dispositions as regards wearing his hat. Bad weather invariably induces him to wear his hat. Fine weather, on the other hand, affects him neither way: on fine days he puts his hat on or leaves it on the peg, completely at random. Suppose, moreover, that actually the weather is bad, so Jones *is* wearing his hat.
>
> (Tichý, 1976), p.271

[69](Lewis, 1979), p.457
[70](Fine, 1975), p.452

We would probably like to say that

(64) If the weather were good, Jones might not be wearing his hat.

is to be regarded as true in the proposed scenario. Yet if Jones's actual wearing of the hat should be counted as a relevant respect of similarity, then the tie between good-weather worlds with Jones wearing and not wearing his hat will be broken in favour of the former, contrary to our intuitions.[71]

Lewis rose to the challenge. He answered his critics by acknowledging the intuitions but denying that they posed a real difficulty for his theory, as the flexibility built in the notion of similarity was able to accommodate them. Such a reply was, of course, slightly disingenuous. We have seen that when selling his theory, Lewis at least intimated that no outlandish similarity relation would be needed to make it work. However, Fine's and Tichý's examples showed that the resemblance criteria standardly governing counterfactuals differed radically from those used in comparisons of everyday objects.

Aware of this fact, Lewis set out to specify the criteria yielding what he called 'the standard resolution of vagueness'. After a thorough analysis of Fine's example, he came up with the following four constraints on our choice of the closest antecedent world:

> 1. It is of the first importance to avoid big, widespread, diverse violations of law.
> 2. It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
> 3. It is of the third importance to avoid even small, localized, simple violations of law.
> 4. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.
>
> (Lewis, 1979), p.472

The basic difficulty anyone can see with this account is its blithely *ad hoc* character. It is reminiscent of F. Bacon's haphazard inquiry into the nature of heat in *Novum Organum*. In the absence of any organising principle, Bacon ended up with an unmanageable list of all possible phenomena

[71] The famous zoo escape example discussed in (Kratzer, 1989), p.126, has approximately the same structure. See also the criticism in (Jackson, 1977).

bearing some relation to it, ranging from burning candles to fresh manure. The bottom line is that proceeding in this manner doesn't bring us any closer to a proper understanding of heat or, for that matter, counterfactuals.

Despite the vagueness of our everyday similarity criteria, we can more often than not easily pin down its source. Sometimes we judge two objects similar, because they can be used interchangeably for a given purpose. Sometimes it is because they share some perceptual marks that are otherwise relevant—for instance, two faces resemble each other if they have traits in common that are crucial in human recognition of other people. Sometimes similarity judgements are meant as warnings about the possibility of confusion of two different things. However, what kind of practical purpose might serve the similarity criteria set forth by Lewis? You might try to argue, following cotenability theoreticians, that they underscore the special status of laws, but it is by no means clear why they should do so in such a convoluted manner.[72]

Moreover, of course, not a few eyebrows were raised at Lewis's reference to miracles. Certainly you don't have to accept unfathomable natural events in order to evaluate counterfactuals? Lewis was perhaps trying to dispel such fears when he observed that the events in question came across as miracles only from the vantage point of the actual world—of course they did conform to the laws of the world in which they took place.[73] However, I for one do not feel reassured by these remarks. We only inhabit the actual world and we evaluate counterfactuals with respect to it—there is no other viewpoint for us to take. I fear that having to postulate zany contexts for consequent evaluation in order to make sense of counterfactuals brings us dangerously close to the bygone attitude that dismissed them as mere exercise in irrelevant fantasising. Lewis's theory once again fails to identify the role counterfactuals play in our intellectual and/or linguistic life.[74]

Lewis, none the less, didn't really seem to mind about this kind of objections, as long as the similarity relation thus calibrated should give the right results. Which, unfortunately, it didn't.

[72] As we have seen (page 63), Lewis roughly shared Ramsey's 1928 conception of laws, under which they amounted to axioms/theorems in the best true deductive systems in a world. Leaving aside the plausibility of such a construal, it cannot straightforwardly explain the constraints 1-4 in terms of some more intuitive respect of similarity, such as match of particular matter of fact, and neither does Lewis attempt an exhaustive explanation along these lines.

[73] (Lewis, 1979), p.468

[74] Similar criticism is voiced in (Woods, 1997), pp.57 and 92.

Take for instance Tichý's example and modify it as follows:[75] upon waking up every morning and before checking the actual weather, Jones flips a coin. If it's bad weather, he puts on his hat as before, but if the weather is good, he only puts on the hat if the coin has come up heads. Today, the coin came up heads, it is bad weather and Jones is wearing the hat. But this time we would like to say that if the weather had been good, Jones would still be wearing the hat. Particular facts (the outcome of flipping the coin) are, all of a sudden, highly relevant for evaluating counterfactuals.

There is a whole host of examples of this kind. They show the importance of causal interconnections between events, which Lewis's theory of similarity fails to bring out.[76] In general, any particular matter of fact not causally dependent on the falsity of the antecedent can become relevant in evaluating a counterfactual:

(65) Even if we had left on time, we would still have missed the train (as there was a spectacular traffic jam in front of the station).

These and related concerns led Stalnaker to admit that the theoretical notion of similarity was to a great extent devoid of content[77]. Lewis, as we have seen, seemed to have favoured an *ad hoc* approach—any constraints whatsoever yielding the correct truth values could be identified with the relevant respects of similarity. The upshot is, again in Stalnaker's words, that

> [...] a formal semantic analysis, by itself, is intended as neither a solution to nor a dismissal of the problem of counterfactual conditionals. [...] It is intended to provide a framework for the formulation of the substantive problems, and for precise statements of alternative solutions to them.
>
> (Stalnaker, 1984), p.122

This is exactly what Stalnaker-Lewis theory has become in modern formal semantics—a scaffolding on an empty construction site. However, it is difficult to put such scaffolding in place before even knowing how the final edifice should look like. It might very well so happen that the pre-installed framework should obstruct rather than aid attempts to come up with novel solutions to the problem of counterfactuals.

[75] This modified example is due to F. Veltman and F. Mulkens, see (Veltman, 2005), p.164.

[76] Actually, Lewis hoped to analyse causality in terms of counterfactual dependence, as yielded by the above constraints on the similarity relation: (Lewis, 1973c).

[77] (Stalnaker, 1984), p.129

4.5. Conclusion: cotenability vs. similarity

To close our analysis of Stalnaker-Lewis account of counterfactuals, I will briefly compare it to a standard cotenability theory. Lewis put forward a definition of cotenability in terms of comparative similarity as follows:[78]

> **Cotenability from similarity**
> $[\![B]\!]$ is cotenable with the counterfactual assumption $[\![A]\!]$ at α iff $[\![B]\!]_\alpha = [\alpha]_R \vee (\exists\beta \in [\![A]\!]_\alpha)\forall\gamma(\gamma \preceq_\alpha \beta \supset \gamma \in [\![B]\!])$[79]

In other words, $[\![B]\!]$ is cotenable with $[\![A]\!]$ at α if B is true in all worlds accessible from α or if some A-world is closer to α than any $\neg B$-world. It is a consequence of this definition that $A \boxdot\!\!\rightarrow C$ is true at α iff there is a collateral premise $\Delta \in \mathcal{P}(K)$ cotenable with $[\![A]\!]$ at α such that $[\![A]\!] \cap \Delta$ logically implies $[\![C]\!]$. Also, if $[\![B]\!]$ is cotenable with $[\![A]\!]$, then $A \boxdot\!\!\rightarrow B$ is true, but not vice versa. The role of laws, an essential ingredient in metalinguistic theories of counterfactuals, is not explicitly distinguished from that of ordinary matters of fact, but of course it is doing its work at the level of the similarity ordering of worlds.

Furthermore, given this definition, we can in turn easily define the similarity ordering from cotenability as follows:

> **Similarity from cotenability**
> $\beta \prec_\alpha \gamma$ iff all propositions cotenable at α with $\{\gamma\}$ are also cotenable with $\{\beta\}$, but not vice versa.[80]

Of course, given the conclusions of the previous section, these results, despite their formal beauty, do not really constitute any progress towards the solution of Goodman's puzzle. In Stalnaker's words, quoted above, similarity only provides a framework for rephrasing this problem, but doesn't offer any help as to how to tackle it. We would first have to refine the relation of similarity in order to get a grip on cotenability.

While Stalnaker-Lewis analysis provided a well-behaved and beautiful logic (probably the main reason behind its success), cotenability theories

[78](Lewis, 1973a), p.57

[79]This definition assumes almost-connectedness of $\preceq_\alpha$. For a generalisation of this definition, see A.2.

[80]This broadly anticipates the strategy that A. Kratzer will use to provide a more adequate grounding for the similarity relation: see Section 5.2.1.

are in a better position to explain why we would want to use counterfactuals in the first place.[81] However, the question as to why the tacit premises in the arguments hinted at by our counterfactuals are so ineffable remains as baffling as ever.

[81] Hence the great success of Kratzer's theory subjunctives, which combines the best part of each of these two approaches (see Section 5.2). Kratzer's theory is also one of the reasons why I think that the arguments in (Fine, 1975) in favour of a principled difference between cotenability and similarity theories are not conclusive.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 5

Beyond similarity

Despite its self-confessed lack of content, Stalnaker-Lewis's theory cast a long shadow. Most theories of conditionals, and especially subjunctives, that came afterwards had to measure themselves against what became a benchmark in the field. They either tried to glean an analysis of the notion of similarity from independent considerations (Kratzer, Lycan, Gauker) or pin down in a non-trivial way how the context for consequent evaluation is determined (Veltman's update semantics as well as causal network theories: Tichý, Schultz, Higgs and others). We shall examine to what extent these analyses can escape the charge of vacuity brought against similarity semantics. Unfortunately, we shall be forced to conclude that the improvement has not been very impressive.

5.1. *Ad hoc* selection: Lycan and Gauker

We shall now have a look at two particularly egregious examples of philosophical theorising, which not only do not attempt to shed any light on how the ingredients necessary for the evaluation of a conditional are retrieved from a context of utterance, but also effectively trivialise the analysis by letting these ingredients fluctuate wildly across contexts. In other words, the selection of relevant alternatives, among which conditionals are called to distinguish and which up to now have been represented by possible worlds, will no longer be guided by unfathomable respects of similarity—even worse, it will be almost completely at the utterer's discretion. Moreover, originally devised with indicatives in mind, these theories do not devote enough attention to their complicated relation with subjunctives, failing to give an acceptable account of the distinction.

5.1.1. Selection by envisaging events: Lycan

Lycan made much of the fact that he replaced possible worlds with events as the cornerstone of his semantic theory, but all in all, this is quite an insignificant change with respect to Stalnaker-Lewis. Events, as Lycan conceives them, are not punctual temporal occurrences, but simply 'circumstances' or partial states of affairs. Any theoretical achievement of such an event theory can thus be effortlessly emulated by possible-world semantics, by effectively treating events as equivalence classes of possible worlds (either explicitly or implicitly within a truth condition clause).[1]

Let us state the truth conditions Lycan attributes to a simple conditional 'If A, C':

Truth conditions conditionals Lycan
C if A iff $(\forall e \in R)(In(e, A) \supset In(e, C))$[2]

Here e ranges over events, $In(e, P)$ means that the proposition P is true of (or in) the event e and R is the 'reference-class' restricting the universal quantifier domain to a specific set of events 'envisaged by the utterer'. The truth condition says that a conditional is true if its consequent is true in all the events envisaged by the speaker in which the antecedent is true.

R functions, in view of the above, just as the smallest antecedent-permitting sphere (making the limit assumption) in Lewis's original formulation of his system[3]. The difference is only that, events being partial states of affairs, there might be envisaged events in which neither the antecedent nor its negation is true, and thus we cannot simply stipulate that a conditional 'If A, C' is true if '$A \supset C$' is true throughout R.[4] However, Lycan's truth condition is the natural generalisation of the latter idea.

Just as Lewis's antecedent-permitting spheres, R can shift considerably between and even within contexts, thus avoiding a strict interpretation of the conditional under evaluation. However, on Lewis's account, these

[1] Nine years before the publication of the first version of Lycan's account in (Lycan, 1984), D. Lewis had already splendidly explained the syntactic phenomena upon which Lycan constructed his theory, showing that they constitute only a marginal fragment in a much broader pattern, in his (Lewis, 1975). However, although (Lycan, 2001) references Lewis's paper, it never addresses the challenge it poses to the event theory. More on adverbs of quantification in section 5.2.

[2] (Lycan, 2001),p.18

[3] See (Lewis, 1973a), p.19. A sphere is simply a set of all possible worlds up to a certain degree of similarity to the actual one.

[4] That is, unless we accept what Lycan calls 'the Strict Relevance Restriction', in which case the conditions coincide completely. Lycan is loath to reject this restriction outright, admitting that it is sometimes triggered in natural language.

shifts are systematically regimented by an underlying system of spheres (or, equivalently, by a similarity ordering) construed as a representation of certain objective criteria of comparison. No such objective foundations are relied upon by Lycan, who allows *R* to vary literally at the utterer's whim, despite affirming that 'the delineation of the reference class is a matter that must be settled fairly crisply before any truly serious adjudication of the theory can be achieved.'[5] Surprisingly, though, he doesn't make any systematic attempt at such a delineation in the whole monograph. The most specific thing he gets to say about *R* is that it perhaps should include all events 'a reasonable person would have envisaged'. Vague though this is, it still doesn't manage to solve the problem, as any reasonable analysis of 'reasonableness' will admit of margins within which reasonable people may envisage different events. We are never told what drives people to envisage one event rather than another. Worse still, we are never told what drives people to envisage events at all.

On Lycan's theory, indicative and subjunctive conditionals differ in that an utterer of the former 'holds fixed a certain contextually determined salient fact', while an utterer of the latter signals that he envisages events beyond such a fact. So, in case of Adams's pair:

(66) If Oswald did not kill Kennedy, someone else did.

(67) If Oswald had not killed Kennedy, no one else would have.

someone asserting (66) envisages only events in which Kennedy is assassinated (the salient fact in this context). When uttering (67), on the other hand, you are making it explicit that you want to consider other events as well. However, considering such new events has the magical effect of rubbing all of the former ones out from *R*, as they would render (67) false. Lycan is silent on how this effect comes about.

And, of course, 'conversationally determined salient fact' is as vague as it possibly could be. Take an example by Mackie[6]: imagine you are in a room with a few people who have had to take a thorough English language test in order to be present. Then you can say both

(68) If Khrushchev were here, not everybody in the room would speak English.

and

(69) If Khrushchev is here, then he speaks English after all.

[5](Lycan, 2001),p.24.

[6](Mackie, 1962),p.71

Following Lycan, a person uttering (69) perseveres in her belief that everybody in the room has passed the test and that the test was reliable. But why should this fact be deemed more salient than the fact that Khrushchev doesn't speak any English, as (68) takes for granted?

All this is not meant to deny that there are valid intuitions behind Lycan's position—it serves rather to point out that his theory boils down to little more than a few casual hunches. The utterly *ad hoc* character of the parameters used for conditional evaluation prevents it from being a *real* account of subjunctives or, for that matter, conditionals in general. I shall conclude with Lycan's criticism of Adams's explanation of the Kennedy examples:

> But the qualification raises questions about how we are to determine what background information we are allowed to and/or forced to count, and a theory of the straight/boxarrow distinction is supposed to answer such questions rather than raising them.
>
> (Lycan, 2001), p.149

I couldn't have put it better myself.

5.1.2. Selection by abstract contexts: Gauker

While Lycan remains blithe about the lack of content of his fundamental definitions and entertainingly triumphant when his impromptu criteria score points against more solid proposals, Gauker doesn't even mind his theory going seriously awry. While most philosophers of language try first to carefully induce their conjectures from linguistic data and subsequently iron out any outstanding wrinkles, for Gauker the predictions of his coarse semantic theory seem to trump most conflicting evidence.

The central theoretical notion in Gauker's definitive theory of conditionals as exposed in (Gauker, 2005) is that of *context*. Here context, in opposition to common practice in semantics, shouldn't be understood as the actual environment of a conversational exchange[7], but rather the formal value of a variable used by the semantic theory to evaluate assertions. In its simplest form, Gauker's context is simply a consistent set of literals. These so-called 'primitive' contexts can also become collected in sets and form

[7]Such environment is rather termed 'situation' or 'circumstance' and seems to correspond roughly to the 'situational elements of the context' from (Gauker, 1998). Situations only partly determine contexts proper.

1st-level multicontexts, the latter can in turn form 2nd-level multicontexts, and so on and so forth.[8]

Unlike Lycan, Gauker severs the context pertinent to a conversation from the particular belief states or envisagings of the participants in it, making it an objective feature of a conversation. The context is somehow jointly determined by both the actual circumstances of the conversation and the shared conversational goals of the participants. An interlocutor's ideas about what the actual context is are represented by his *take* on the latter, which may or may not coincide with or, for that matter, include the context really pertaining to the conversation. However, the price for this separation between the objective context and a subjective take on it is that the former notion becomes pretty elusive, especially since there is no straightforward way of gleaning it from the circumstances and conversational goals.

The function of an assertion is similar to what a possible-world semantics along Stalnaker's lines might posit: to narrow down each interlocutor's take on the context. Gauker also replaces truth *simpliciter* with *assertibility in context* as the linchpin of his semantic analysis. A literal is assertible in a primitive context only if it is a member of such context, and it is deniable if its negation is assertible. A literal is assertible in a multicontext only if it's assertible in each of the lower-level contexts that comprise it. Logically compound sentences follow, for the most part, the expected evaluation rules.

Let us now turn to conditionals. Here are Gauker's definitions:

> *Indicative conditionals*
> $A \rightarrow C$ is assertible in a context Γ if in all contexts $\Delta \in \Gamma$ or $\Delta = \Gamma$ such that A is assertible in Δ, C is also assertible in Δ.
>
> *Subjunctive conditionals*
> $A > C$ is assertible in context Γ_0 with respect to a structure Θ, where Θ contains the context immediately relevant to the conversation Γ_0 as well as its less relevant expansions, if $A \rightarrow C$ is assertible in all least expansions of Γ_0 that include at least one lower-level context in which A is assertible.

In case of indicatives with non-conditional operands (I will address nested conditionals later), contexts thus conceived serve a purpose by and large akin to Lycan's reference-classes built out of partial states of affairs

[8](Gauker, 2005), pp.12-24. A higher-lever context can also contain elements corresponding to different levels.

and as such can be roughly adequately rendered in possible-word semantics (despite Gauker's objections). We can again represent a primitive context as a partial world (defined, say, as an equivalence class of possible worlds). Then a 1st-level multicontext will amount to a set of such partial worlds, just as Lycan's reference-class. In this particular case, Lycan's and Gauker's evaluation conditions coincide exactly.[9] And while this time the set of relevant possibilities is defined independently from the mind of the utterer (which is undeniable progress with respect to Lycan), we still have no independent grip on how this set is constructed.

One difference between the proposals is that for Lycan the utterance of a conditional forces the speaker to envisage at least one event in which the antecedent is true. Gauker doesn't accept this constraint, as his definition of validity in terms of assertibility-preservation requires contexts to remain unchanged from premises to conclusion, whatever form they may have. This leads him to accept that indicative conditionals are *vacuously* assertible in a context whenever the antecedent is not assertible in any of the subcontexts, with all the difficulties that such a position brings in its wake.

Gauker's assertibility condition for subjunctives is perhaps less intuitively clear. The basic idea is that when an indicative conditional is only *vacuously* assertible in a context, we can sometimes find larger contexts, 'less relevant' to our present purposes, which do contain a subcontext sustaining the antecedent. A subjunctive is assertible in the original context if the corresponding indicative is assertible in all the minimally augmented ones.

We have seen that Lycan's explanation of the indicative/subjunctive distinction also involved expansion of the reference class beyond a certain salient fact to which an utterer of the indicative held firmly; the problem being that this relaxation led to an inexplicable erasure of certain events from the originally envisaged *R*. Gauker doesn't have this problem—in case of a difference in truth values, such as with the Adams's pair (66) and (67), the original context is by hypothesis devoid of antecedent-subcontexts, so there are no annoying subcontexts to shed. However, this comes at a price. For Gauker, when (67) is assertible, then not only (66) is, but also

(66*) If Oswald didn't kill Kennedy, then no one else did.

as both (66) and (66*) will be vacuously true.

[9] This may serve as a reminder that simply substituting 'assertibility' for 'truth' or vice versa alone doesn't necessarily make for much difference in a semantic theory.

By now it will have dawned on a keen reader that the difference between Gauker and Lycan (again, leaving embedded conditionals aside) is merely superficial. Gauker will not, of course, deny that we tend to regard (66) as non-vacuously true, but he will posit a context shift between the utterances of (66) and (67). Just as with Lycan, context-shifts are going to occur whenever Gauker needs them to vindicate his theory.

Surprisingly enough, this brings Gauker's theory closer to classical similarity semantics than he might be willing to concede. Everybody acknowledges the presence of some kind of contextual shifts in non-monotonic conditional reasoning; the really interesting question is to explain why they take place and, if possible, spell out the rules that govern them. Similarity semantics is founded on the presumption that the shifts are triggered by antecedents and regimented by a contrived similarity relation. This is admittedly not to say much, but it is still more interesting than what Gauker has to say about the shifts (virtually nothing). For instance, if there is a shift between (67) and (66), why don't we switch to one of those minimally expanded antecedent-permitting contexts in the alleged Θ?

This way of looking at Gauker's theory also takes the edge off the logic ensuing from his definitions, which sets his conditional apart as a peculiar blend of the material and strict conditional. Gauker's account validates most traditional inference patterns that Stalnaker-Lewis theory was specifically designed to avoid. However, this is only because of Gauker's insistent prohibition of contextual shifts. If we took them into account, the difference in logic would arguably be much less strident. Again, logical questions frequently depend on the phenomena we aim to model, which gives us quite some room for manoeuvre. Postulation of divergent logics thus doesn't necessarily imply any profound contrast in the underlying theoretical approach.

This is borne out by the disparity between indicative and subjunctive logic embraced by Gauker. It is due to the fact that Θ provides a model for contextual shifts and, as a result, a basis for a standard similarity ordering (however, the notion of a 'smallest expansion' would first need to be made more precise). The analogous indicative shifts are, none the less, simply left unaccounted for by the theory—thence the diverging logic. All in all, there are good reasons to think that conditionals are conventional devices whose task is to accomplish shifts of our take on the context, however they may further be construed, so banishing them as a matter of principle from the realm of conditional logic may be unfortunate.

Of course, Gauker would undoubtedly object to such an oversimplified rendering of his theory, and he would probably be right. In particular, his theory can elegantly handle conditionals embedded to any degree of

depth by means of a nested system of contexts. This complicates the logic and sometimes gives rise to doubts whether the results accord with our intuitions[10], but on the whole Gauker's system is on this point much more flexible than competing semantic theories.

However, this flexibility is obtained only thanks to the fact that the contexts are always and on every stage perfectly geared for the particular sentence under evaluation. Let us begin with primitive contexts: as we do not know how they are generated, they do not offer an independent handle on the notion of assertibility—literals are not assertible because they are in the context set, but they are in the context set, because they are assertible.

According to Gauker, multicontexts make their entrance for example when indecision, unpredictability, didactic considerations or even fairness is involved. It is by no means clear how this can be accommodated with the purported mind-independence of the contexts. But more importantly, we still do not have any grip on the constraints on the generation of such complex contexts. Very likely it is our intuitions about assertibility of conditionals that guide this process. Moreover, these allegedly objective contexts can sometimes split for different people.[11]

When we want to assert embedded conditionals, nested contexts again appear out of the blue, giving us just what we need to evaluate them. And the same casual approach also characterises the introduction of the system of contexts Θ that governs the evaluation of subjunctives. The system is *ad hoc* all the way up.

5.1.3. Instead of a conclusion

Lycan and Gauker's theories are only illustrations of a broader trend in some of the literature on conditionals. Unfortunately, you can go a long way without saying anything substantial or at least verifiable about the topic. And once you're done, you can always start comparing the formal properties of different nearly-vacuous theories. This is a huge boon for many a scholar desperate to get another paper published; for the rest of mortals it boils down to an enormous waste of time, effort and, more often than not, public resources.

[10] My reservations concern above all the interpretation of 'weak validity', redolent of Lycan's various restrictions that were triggered *ad hoc* in conversational settings. Moreover, some intuitively acceptable inference patterns come out invalid on Gauker's semantics. An example is *And-If*, called 'exportation' in (Gauker, 2005), p.121: $(A \wedge B) \rightarrow C \models A \rightarrow (B \rightarrow C)$. Gauker labels it as valid, but I think that's due to oversight.

[11] See Gauker's treatment of Gibbard's puzzle: (Gauker, 2005), p.103.

5.2. Premise Semantics

In his (Lewis, 1975), D. Lewis pioneered a stunningly new approach to conditional sentences. According to this theory, certain if-clauses function exclusively as domain restrictors of (possibly implicit) natural language quantifiers. Thus in a sentence such as

(70) Always, if a man owns a donkey, he beats it now and then.

we have an adverb 'always' quantifying over cases (conceived as ordered pairs of objects) and an if-clause restricting the domain of quantification to the relation of donkey owners and donkeys. Lewis argued that such ifs couldn't be plausibly equated to any binary sentence operator, barring a considerable overhaul of their standard logic. He concluded that such if-clauses didn't have any meaning apart from their task in restricting the adverb.

Lewis preferred to embrace polysemy rather than bite the bullet and explore the theoretical opportunities his novel insight offered for the study of conditionals. He seems to have overestimated the concomitant danger for the role classical logic and its material conditional played in the semantics of indicatives, to which he was firmly wedded. It was thus only one of Lewis's most talented students, Angelika Kratzer, who rose to the challenge of extending the promising restrictor approach to all conditionals. Even though I will raise some objections to this approach in what follows, Kratzer's work proved impressively fruitful and set a new paradigm in natural language semantics.

5.2.1. Kratzer on modals

Kratzer's theory of conditionals is an application of Lewis's insight about restrictive if-clauses to her own theory of modals as peculiar quantifiers over possible worlds. The main theoretical goals of Kratzer's approach to modals comprised an explanation[12]

i) of the perceived polysemy between their epistemic, deontic, and other uses

ii) of our ability to reason non-trivially from inconsistent sets of premises.

The first problem was addressed by putting forward a relational theory of modals. According to Kratzer, each modal expresses a certain relation

[12](Kratzer, 1977), pp.9-16

between the *modal scope* (that is, the proposition filling in its explicit argument place) and a pragmatically determined (and often implicit) modal restriction or *premise set* (a set of propositions against which the modal scope is evaluated).[13] So the perceived difference in meaning between

(71) He must finish the work by one o'clock, or he gets fired.

and

(72) He must (have) finish(ed) the work by one o'clock, it's pretty easy.

is due to a different value (epistemic and deontic, respectively) the modal restriction takes in each case, rather than to a plurality of core meanings of 'must'. Resorting to pragmatics is quite plausible here, as Kratzer points out that the posited elliptic element can usually be brought to the fore by an explicit phrase beginning with 'in view of'.[14]

In the most simple cases, 'musts' will be deemed true if the modal scope logically follows from the premise set provided by the modal restriction, and 'cans' (or, for that matter, 'mays') will be true if the modal scope is compatible with it. From the point of view of the classical modal logic, a consistent premise set therefore yields the set of worlds accessible from the actual one by the relation R that is used to interpret the standard modal operators.

Formally, Kratzer defines a function $f : K \to \mathcal{P}(\mathcal{P}(K))$, called *conversational background*, which assigns a premise set to each possible world.[15] We now obtain the following formal rendering of the above truth condition:

TC musts Kratzer formal
$[\![\text{Must } A]\!]^f = \{k \in K \mid (\bigcap f(k) \subseteq [\![A]\!])\}$

and for can/may

TC cans Kratzer formal
$[\![\text{Can } A]\!]^f = \{k \in K \mid \bigcap(f(k) \cup \{[\![A]\!]\}) \neq \emptyset)\}$.

[13] Hence the term 'premise semantics', actually coined by Lewis in (Lewis, 1981).

[14] (Kratzer, 1977), p.6

[15] This is only to mean that premise sets can vary across different situations—a conversational background should be viewed as assigning one such set to each conversation within which the modalised statement is asserted. By setting K as the domain of f, Kratzer can make modals and conditionals work within sentences expressing traditional propositions. However, we will presently see that there are good reasons to require a more fine-grained domain for f.

These truth conditions can be further complicated in order to account for graded possibility. Sometimes, one wants to order the different alternatives compatible with a given premise set according to a certain criterion, such as likelihood or closeness to some ideal. That's why you can say things like

(73) The cook as well as the maid may have murdered the postman. However, the maid is more likely to be the murderer, as she had had a long-standing feud with him.

Another, less obvious application of graded modality is provided by deontic interpretations:

(74) You must return the stolen toothbrush to the store.

is true if returning the toothbrush is the course of action most compatible with a certain moral ideal (the fact that graded modality is needed here follows from the fact that none of the live possibilities is compatible with the ideal as a whole—the toothbrush cannot simply get unstolen.)

Kratzer treats such cases by resorting to two different conversational backgrounds:[16]

1. *Modal base f:* as before, it delimits the entire range of relevant possibilities in a given context of evaluation.

2. *Ordering source g:* imposes an ordering upon the set of possible worlds admitted by the modal base.

I will sometimes refer to the set of worlds not ruled out by the premise set $f(k)$ as f-accessible worlds for k. An ordering source $g(k)$ being a premise set, that is, a set of propositions, the intended ordering is defined by positing that a world i is 'at least as possible as' j with respect to $g(k)$ if it supports all propositions from $g(k)$ that j supports and possibly more.

Ordering induced by $g(k)$
For $i, j, k \in K$, $i \leq_{g(k)} j \equiv_{def} \{p \mid p \in g(k) \wedge j \in p\} \subseteq \{p \mid p \in g(k) \wedge i \in p\}$

This ordering will in general permit both ties and incomparabilities, so it might look like a partial order, but the failure of antisymmetry means that it is only guaranteed to be a preorder. However, a partial order can of course be defined in terms of it so as to match the most general type of

[16](Kratzer, 1981b)

similarity ordering.[17] Kratzer's self-acknowledged goal was to reconstruct the similarity logic for counterfactuals with theoretical building blocks better-suited for the task.[18]

According to Kratzer, for the deontic necessity from the example (74) to be true it suffices that you wind up returning the toothbrush in all the worlds *closest* to the ideal. For sheer possibility the ordering may just as well be ignored; for likelihood some intermediate criterion can be devised. The truth condition for each modal will vary accordingly. Kratzer takes it upon herself to show how the semantic import of different natural language modalities can be rendered in this framework.

5.2.2. The restrictor approach: indicatives

To combine the above account of modals with Lewis's treatment of restrictive if-clauses, two preparatory steps are called for. First, we have to realise that Kratzer's approach enables us to conceive modals along the lines of Lewis's adverbs of quantification, as quantifying over the set of possibilities not ruled out by an often tacit premise set supplied by a modal base. And second, we have to posit an elliptic modal in the consequent of many conditionals with apparently non-restrictive if-clauses, even if on the face of it, such conditionals don't contain any modal vocabulary. This is not utterly implausible, because most of these cases will be of universal quantification, which is often left implicit in non-modal discourse as well. Once these two steps are carried out, everything falls into place.

Let us begin with the more simple case of indicatives. Kratzer contends that any bare conditional such as

(75) If the lights in his study are on, Roger is home.

should be regarded as implicitly modalised:

(75m) If the lights in his study are on, Roger *must* be at home.[19]

Now you can apply Lewis's insight in a straightforward way. Your premise set (consisting, say, of propositions like 'Nobody but Roger uses the study', 'Lights are on only if there are people in the house', etc.) carves out a subset of K that serves as a domain of quantification for the tacit modal. Supposing for the sake of the example that there is no need for

[17] See Section 4.2.5

[18] (Kratzer, 1989), p.126

[19] (Kratzer, 1991), p.98

grading the modal force and thus for an ordering source, the intended restriction can then be accomplished by simply enriching the above premise set with the proposition expressed by the antecedent:

TC if-must Kratzer
$[\![\text{If } A, \text{ must } C]\!]^{f,g} = \{k \in K \mid f(k) \cup [\![A]\!] \subseteq [\![C]\!]\}$.

In general, a modal statement with a conditional restriction will be true whenever the consequent comes out true with respect to the pertinent premise set enriched with the proposition corresponding to the antecedent.

General TC ifs Kratzer
$[\![\text{If } A, C]\!]^{f,g} = [\![C]\!]^{f^*,g}$,
where $f^*(k) = f(k) \cup [\![A]\!]^{f,g}$ for all $k \in K$.[20]

But can this account really provide a general treatment for all indicative conditionals? I will argue that there are a few difficulties that must be addressed in order to make it work.

First, I am yet to be convinced that all bare indicatives are implicitly modalised, unless we stretch our notion of modality beyond recognition. Admittedly, many conditionals sound equivalent to their modalised counterparts, but it is by far not always the case. None of the following examples admit of a modal in their consequent:

(76) If that was a police siren, we're in trouble.

(77) If he thinks anybody cares about his PhD, he is in for a surprise.

(78) If she finally broke up with him, she did the right thing.

A special case are conditionals conveying promises, pledges, threats, etc., whose consequent usually contains the first-person future form of the main verb.

(79) If you do not want your tripe soup, I will be glad to have it.

(80) If he doesn't stop whistling immediately, I will have him thrown out of here.

The above conditionals and their kin do not admit of modalising with a 'must', but one can convincingly argue that 'will' should be viewed as a modal in its own right. However, this leaves open the question about what kind of modal it is. And the fact that many languages (such as Romance

[20](Kratzer, 1981b), p.65

languages and most Slavic ones) feature simple future tense casts a shadow of doubt on the plausibility of this way out of the difficulty.

Kratzer claims that when bare conditionals are preferred to their modalised counterparts, it is because they are asserted 'without relying on any particular piece of evidence at all', which makes a bet on them into a bet on the corresponding material conditional.[21]. This doesn't sound right—intuitively, a conditional bet doesn't win when the antecedent comes out false. Moreover, if a bet on *C* winds up losing, so must a bet on *must C*, regardless of the evidence one might have had for the latter at some earlier point.

Be it as it may, there is further evidence against implicit modals in the consequent, such as conditional advice, exhortations or orders, which lack truth conditions anyway:

(81) Should you experience any difficulties, don't bother to call the help desk—they're useless.

And, finally, some antecedents seem to place a condition on the act of utterance rather than on the content of the consequent:[22]

(82) If I may interrupt, there is no such thing as the greatest ordinal.

My second reservation about Kratzer's treatment of indicatives has to do with the fact that it bestows traditional truth conditions on indicatives. Does it mean that we have found a way to circumvent Edgington's arguments after all? What about the Equation, does it still hold in Kratzer's framework? And does that make the theory liable to Lewis-like triviality?

The short answer to the first question is no. Recall how we determined the truth value of a simple necessity conditional in a possible world k: we first applied f in order to obtain the premise set $f(k)$ corresponding to the modal base, we enriched the latter with the proposition expressed by the antecedent, and we checked whether the thus restricted domain implied the consequent. The modal base for any k is usually realistic, that is, all its propositions are true in k, but it doesn't have to be *totally* realistic, that is, it needn't identify k uniquely. So we end up with small fixed spheres of uncertainty associated with each world—and that's why we can non-trivially evaluate an indicative across the whole K.

But how plausible is it to attach one (or two, in case an ordering source is needed) particular premise set to a world as a whole? Not very, I submit. Intuitively, particular premise sets characterise individual exchanges, or

[21](Kratzer, 1991), p.99

[22]See also (Bach, 1999), pp.356-358.

perhaps only fragments of them, myriads of which inhabit each possible world. If, on the other hand, we take a probability space to represent the accessible possibilities at a given point of a conversation, then it will also represent the relevant premise set for simple necessity conditionals—we do not have to posit one for each crisp possibility the speaker recognises.

We are compelled to conclude that there is less to the distance between Kratzer and Edgington than meets the eye. Viewed from this perspective, Kratzer's proof of the contention that indicatives degenerate to material conditionals if evaluated against totally realistic modal bases[23] is essentially a variant on Edgington's own argument against truth conditions for indicatives.[24] It suggests that if all actual facts are taken into account, there is no way to escape the material interpretation of indicatives. The upshot is that for both authors, not possible worlds but premise sets determining speaker's live possibilities are the correct points for evaluating indicative conditionals. Assigning one such set to each world just to safeguard the traditional semantic framework has an artificial whiff to it while being theoretically ineffectual.[25]

In view of the foregoing, it should be clear that strictly speaking, there is no guarantee for Kratzer's indicatives to satisfy the Equation.[26] Unless you somehow constrain the choice of the modal base for the worlds in K, the truth value of an indicative at each world will for all intents and purposes be independent of those of its operands. Therefore, its absolute probability can diverge wildly from the conditional probability of the consequent given the antecedent.

The situation of course becomes different at the level of individual worlds with fixed associated premise sets. We can conceive of the set of worlds f-accessible from k (that is, $\bigcap f(k)$) as themselves defining a probability space. Egré and Cozic have convincingly argued if you interpret probability assignments as proportional quantifiers, then such assignments to conditional sentences should, given the restrictor approach, match the ratio formula.[27]

[23](Kratzer, 1981b), pp.65-66

[24]See Section 2.3.2.

[25]To be fair, Kratzer eventually concurred ((Kratzer, 1991), p.99), admitting that circumstances of evaluation must be more fine-grained than possible worlds. This brings her even further apart from traditional truth-conditional semantics. The question remains, though, as to how these circumstances are individuated and by which mechanism they determine the correct premise sets. The frequent indeterminacy of truth conditions for modals on this account may have to do more with this theoretical loophole than with genuine modal underspecification.

[26]See Section 2.2.

[27](Egré and Cozic, 2011), pp.20-21

This, however, clashes a little with Kratzer's own treatment of bare conditionals of the form 'if A, C', according to which C expresses an implicit necessity, and thus the conditional must be assigned zero probability whenever $p(C/A) < 1$. This is another reason to be suspicious of freely positing implicit necessities. One way out would consist in plugging in an ordering source, but this doesn't seem to be compatible with the assumption (which Egré and Cozic make) of equiprobability of all worlds in $\bigcap f(k)$. The best option would arguably be to switch from 'must' to 'highly probably' as the posited implicit modal (you can, but don't have to, subsequently resort to an ordering source). But then, couldn't 'highly probably' accompany most of our non-conditional assertions as well? In order to responsibly assert a proposition, you should be pretty sure that it is true, even though everybody knows that complete certainty is hard to come by. If the same happens with conditionals, there doesn't seem to be any need, apart from sheer theoretical convenience, for postulating any tacit modals in the consequent.

There is another related difficulty with Kratzer's employment of ordering sources for indicatives coupled with the default necessity reading of bare conditionals. Kratzer needs non-empty ordering sources in order to account for the frequent failure of inference patterns such as AS, CP, or HS, valid on the material and strict interpretation.[28] But limiting the $f-$accessible worlds relevant for the interpretation of a 'must' to the g-closest ones yields counterintuitive results. If you have $A \rightarrow must\ C$ and $(A \wedge B) \rightarrow may\ \neg C$ (note, by the way, that these two alone already sound contradictory) you will in general also have $A \rightarrow must\ \neg B$. This combination sounds all right both for bare indicatives and for counterfactuals, but arguably leaves a strange aftertaste with *must* and *may* in the consequent.

Be it as it may, we are essentially back to Adams-Edgington's position, according to which indicatives are not evaluated at each point of $\bigcap f(k)$, wherefore they fail to divide it into two disjoint regions, as one would expect from a 'genuine proposition'. That's why standard probability laws do not apply to conditional sentences, undercutting Lewis's triviality proofs. In particular, Egré and Cozic maintain that it is the law of expansion by cases (EbC)[29] that should be viewed with suspicion when applied to conditionals. If if-clauses' semantic contribution consists exclusively in restricting quantifier domains, thus in particular equating probabilities of conditionals with conditional probabilities, the following principle, on which (EbC) fundamentally depends, must fail:

[28](Kratzer, 1981b), p.68

[29]See page 40.

(EbC$^\wedge$) $p(A \rightarrow C) = p((A \rightarrow C) \wedge B) + p((A \rightarrow C) \wedge \neg B)$,

since there is no principled way to assign probabilities to compound formulas such as $(A \rightarrow C) \wedge B$, exactly as was the case in Adams's system.[30]

All in all, I do not think Kratzer's treatment of indicatives enjoys much edge over Edgington's and Adams's approach. Modal restriction by conditional clauses is a genuine phenomenon and Kratzer deserves all the credit for having drawn theoretical attention to it. I doubt, however, that it should be the last word on indicatives. In particular, I hope to have shown that there are good reasons to be sceptical about the magnitude of modal ellipse which the theory requires. I suggest that we should regard modal restriction as just one of many important tasks that conditionals are capable of performing in virtue of their more general meaning. However, before jumping to conclusions, let us see how Kratzer's theory fares when applied to subjunctives.

5.2.3. The restrictor approach: subjunctives

The general theory of modals, employing both modal bases and ordering sources, was devised by Kratzer to account among other things for reasoning from inconsistent sets of premises. And this is, of course, just what we need for the purpose of evaluating counterfactuals, which make assertions under assumptions contradicting what is really the case. In this way, Kratzer managed to integrate the intuitions behind both cotenability and similarity approaches to counterfactuals.

Kratzer's basic recipe for evaluating counterfactuals, inspired by cotenability theorists, is to combine the antecedent with some actually true additional premises and check whether the consequent follows. However, by now we know that there are usually several different ways to accommodate the antecedent to reality while preserving consistency—according to Kratzer, we must pay equal heed to all of them. Thus, informally, $A \boxdot\!\!\rightarrow C$ will be true if all consistent ways of supplementing A with actually true premises will yield C.

Within Kratzer's framework, this result can be accomplished by construing 'would' and 'might' respectively as a necessity and a possibility modal, and interpreting them with respect to an empty modal base and a realistic ordering source. The ordering source will provide a set of true premises which will induce a relation on A-worlds that could justifiably be regarded as a typical similarity ordering—if $i \leq_{g(k)} j$, then i supports more

[30](Egré and Cozic, 2011), p.25. For comparison, see (Adams, 1975), p.35.

premises from $g(k)$ than j, and therefore should intuitively be deemed closer to actuality. Here, it is the reality that plays a role analogous to the 'moral ideal' used to interpret example (74).[31]

A would-counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ will come out true on this semantics if $[\![C]\!]$ follows from all the maximal consistent sets containing $[\![A]\!]$ together with propositions from the premise set supplied by the ordering source. However, even if under normal circumstances we can rely on the existence of such maximal sets, Kratzer wishes to allow for the possibility that you could go on enriching $[\![A]\!]$ with propositions supplied by the ordering source indefinitely without eventually arriving at an inconsistent set. Taking this into account, let us define X_Γ as the set of consistent subsets of Γ.[32] Then $A \mathrel{\Box\!\!\rightarrow} C$ will come out true if for any consistent way of combining true premises with the antecedent, you just have to add some more so that the whole imply the consequent.

For Kratzer, as for Lewis, might-counterfactuals are dual to the would-ones. In other words, a might-counterfactual $A \mathrel{\Diamond\!\!\rightarrow} C$ is true if there is a consistent subset of the enriched conversational background containing $[\![A]\!]$, all of whose supersets are compatible with $[\![C]\!]$, or simply if there is a maximally consistent subset of $g(k) \cup \{[\![A]\!]\}$ containing $[\![A]\!]$ and compatible with $[\![C]\!]$.[33]

Those who have carefully read the chapter on cotenability theories know that this cannot be the last word on the subject. Goodman was also on the lookout for the right antecedent-compatible subset of all facts, entailing the consequent if and only if the counterfactual was true. His take on counterfactuals failed precisely because there usually were antecedent-compatible subsets of facts that produced wrong results.[34] If Kratzer let them in, they would wreak havoc on her analysis, so g cannot in general be a totally realistic conversational background. This compels us to specify how the premise set $g(k)$ is selected for each k. The very same difficulties that beset Goodman's account haunt a naïve version of premise semantics as well.

To her credit, Kratzer never subscribed to a naïve version of premise semantics for counterfactuals, with totally realistic ordering sources. From the beginning, she was aware that the relevant premise set had to be conveniently restricted and structured so as to yield the right outcome.[35]

---

[31] See page 117.

[32] A set of propositions Θ is consistent iff $\bigcap \Theta \neq \emptyset$.

[33] For a precise formal rendering of Kratzer's truth conditions for would- and might-counterfactuals, see A.3.

[34] See page 58.

[35] See for instance (Kratzer, 1977), pp.16-20 and (Kratzer, 1979), p.123.

The challenge was to supply an adequate theoretical treatment of such restrictions.

5.2.4. Enter lumping

To be sure, in order to avoid the difficulties of the naïve version of premise semantics, one can simply calibrate the relative weights of different facts by plugging in an ordering source as codified by a Lewis's system of spheres.[36] This is also what Kratzer originally proposed in (Kratzer, 1979). However, seeking both more generality and an independent rationale for the similarity ordering, she eventually addressed the issue by bringing into play the notion of *lumping*.[37]

Let us see what happens if we let everything that is the case in the ordering premise set, so that $g(k) = \{p \in \mathcal{P}(K) \mid k \in p\}$ for all k. Let A be an assumption counterfactual in the world α and C any sentence logically compatible with A. Then $\alpha \in [\![\neg A \vee C]\!]$. Since C is compatible with A, $\{[\![\neg A \vee C]\!], [\![A]\!]\}$ will be consistent and hence eligible for membership in $X_{g(\alpha)\cup\{[\![A]\!]\}}$. Moreover, it implies C, and so do of course all its consistent supersets. The conclusion is that whenever A is consistent with C, $A \diamond\!\!\rightarrow C$ will be true in any $\neg A$-world. The upshot for would-counterfactuals is that unless A strictly implies C, $A \mathrel{\Box\!\!\rightarrow} C$ will come out false in all $\neg A$-worlds, because of A's compatibility with $\neg C$ and the ensuing truth of $A \diamond\!\!\rightarrow \neg C$ (incompatible, in turn, with that of $A \mathrel{\Box\!\!\rightarrow} C$).[38] As we typically use subjunctives as counterfactuals, this would be a devastating problem for the analysis.

Kratzer's solution will consist in banning $\neg A \vee C$ and like 'artificial facts' from $g(\alpha)$. The intuition guiding the search for a more suitable treatment suggests that facts do not come and go alone, but rather in clusters, or 'lumps'. Thus in the above example, the truth of $\neg A \vee C$ in α depends crucially on the fact that $\neg A$, so they should stand or fall together. In Kratzer's idiom, '$\neg A \vee C$ *lumps* in α the fact that $\neg A$'. If the former always brings the latter in its wake, there is no longer any way of constructing a consistent premise set eligible for membership in $X_{g(\alpha)\cup\{[\![A]\!]\}}$ containing both

[36] See footnote 3. As each sphere is nothing more than a set of possible worlds, that is, a proposition, a system of spheres is a suitable candidate for a premise set.

[37] See (Kratzer, 1981a).

[38] This is exactly the objection Parry addressed to Goodman's truth condition in (Parry, 1957), p.86. Although Goodman didn't consider might-counterfactuals, his negative clause has exactly the same effect on woulds as Kratzer's universal quantification over premise sets, (almost) consigning them to the status of the strict implication. See also Kratzer's slightly different proof in (Kratzer, 1981a), p.202.

$\neg A \vee C$ and A.

So the task of an ordering source will consist in selecting the right set of non-trivial facts for a world k. It may still contain propositions that determine k uniquely, but it will no longer have to encompass all propositions true of it.[39] In the case we have been considering, one of the propositions will be $[\![\neg A \vee C]\!] \cap [\![\neg A]\!] = [\![\neg A]\!]$. It might of course yet turn out that $\neg A$ itself lumps other propositions and thus cannot act as an independent fact, but this degree of granularity is sufficient to ward off the difficulties we have observed.

Lewis has shown that if the only constraint imposed on the partition function f assigning a conversational background to each world is that it must determine the world uniquely (a requirement equivalent to strong centring), then Kratzer's semantics are formally equivalent to the infinite generalisation of Pollock's similarity semantics admitting both ties and incomparabilities in the similarity ordering—that is, to the most general similarity semantics we have presented here.[40]

So Kratzer offers us a viable alternative to similarity semantics, which emulates its logical strength, but rests upon independent theoretical underpinnings. And lumping is a real phenomenon upon which we have a sound intuitive grip—in Kratzer's example, to paint a still life involves painting all the flowers or fruits that make it up.[41] Analogously, to return to a difficulty with Goodman's analysis, my not being in North Korea is part and parcel of my not being in Korea, which in turn is lumped by me actually being in the south of Spain. Therefore, adding the counterfactual hypothesis 'I am in Korea now' to the ordering premise set has the effect of eliminating all the above propositions in one fell swoop on pain of inconsistency. By admitting lumping as a cornerstone of our analysis, we can thus prevent many problems besetting Goodman.

However, we must take care not to overreach ourselves. First, the intuitions behind lumping, though strong, are still in need of pinning down. Does my breathing lump my heart's beating? And what about the growth of my hair? Kratzer is willing to cut us some slack in how we can carve up our world, accounting conveniently for the perceived vagueness of counterfactuals[42], but I hope that we have by now grown wary of this kind of hand-waving. We need at least a rough and ready idea of how the context impinges on the criteria governing lumping in order to understand the role lumping plays in our cognitive and linguistic lives. Lumping, just

[39](Kratzer, 1981a), p.211
[40](Lewis, 1981), p.232
[41](Kratzer, 1989), p.112.
[42](Kratzer, 1981a), p.211

as ordering worlds according to some similarity criteria, shouldn't be just 'something we do'—as we presumably take it to be a rational practice, we should be able to gesture towards some purpose that it serves.

Second, for partition semantics to work, lumping has to be more pervasive than its informal presentation might suggest. To see this, let us have a look at how it handles Goodman's problem. Recall that it consisted in explaining why (42) is so clearly preferable to (42*):

(42) If the match had been struck, it would have lighted.

(42*) If the match had been struck, it wouldn't have been dry.

On Kratzer's simple partition semantics, this result is obtained by positing that the match's not lighting lumps its not being scratched. So after adding the antecedent of (42) to the set of facts supplied by the ordering source, the fact that the match actually didn't light is no longer available for any premise set, while the fact that it was dry stands unscathed. However, in the absence of a more solid theoretical underpinning, the tenet that 'not-lighting' lumps 'not-being-scratched' sounds more than a little *ad hoc*. In general, for this semantics to work for counterfactuals, a fact will have to lump its causal antecedents at least up to the negation of the antecedent. Moreover, in counterfactuals that are not obviously strictly causal, lumping will produce quite strange bedfellows:

(83) If you had dropped by yesterday, I wouldn't have made dill sauce.

That your not visiting your friend yesterday should be linked with your having stayed instead at your place watching football sounds quite reasonable, but lumping it with the dill sauce dish seems to be motivated exclusively by the desired outcome for the counterfactual at hand.[43]

Kratzer was well aware of these difficulties and she endeavoured to address them in her subsequent work. Incidentally, this brought her even closer to classical cotenability theories, as she was forced to award a special role to lawlike generalisations. In the simple version of lumping semantics, non-accidental generalisations were assumed to implicitly guide the way the world is partitioned into facts, which left an undesirable *ad hoc* aftertaste. For this reason, more light had to be shed on their lumping behaviour.

[43] See also the criticism in (Tichý, 1984), pp.162-3.

5.2.5. Premises and laws

I do not think Kratzer's later work is correctly interpreted as superseding partition semantics, as some have suggested.[44] It is much more likely that subsequent developments were aimed at fleshing out the sketchy framework of (Kratzer, 1981a) and delving deeper into the central notion of lumping.

Situation semantics put forward in (Kratzer, 1989) constitutes an attempt to formally define the lumping relations that ultimately determine the partition function f. Without going into details, a proposition p lumps the proposition q in the world k if and only if $k \in p$ and all situations (construed as fragments) of k that support p also support q. However, this amounts to little more than a paraphrase of the original idea that lumped facts are somehow inseverable from lumpers. The real merit of situation semantics lies in their ability to bring out the special role non-accidental generalisations play in counterfactual reasoning.

For Kratzer, an accidental and non-accidental interpretation of a generalisation are logically equivalent, but their difference is semantically represented at the level of their lumping properties. While accidental generalisations are relatively strong lumpers and therefore can easily get knocked out of a premise set, lawlike propositions, if true in a world, are true in all its situations and therefore are lumped by all the other true propositions. Note, incidentally, that this is not an account of *why* some generalisations are assigned an accidental interpretation and others not, and which ones it should be at that. That we can successfully make the difference is presupposed all the way—the theory only endeavours to provide an adequate semantic representation of this difference that would

[44] See (Kanazawa et al., 2005), p.135. The authors argue that Lewis's equivalence result showed that Kratzer's theory was open to objections to similarity semantics along the lines of Tichý's Jones example and that this brought Kratzer to revise the theory. However, this can't be—lumping was in one or other form present in Kratzer's theory from the start, and Lewis's equivalence proofs depend crucially on this fact. In presenting lumping semantics as an improvement on partition semantics, Kanazawa et al. even fail to mention that the notion of lumping was explicitly introduced together with the latter. What is more, Kratzer's variant of Tichý's example featuring a zebra escaping from a zoo is devised to show that lumping is up to the job of correctly calibrating the similarity ordering (this turns out to be much less clear with Tichý's original example). Tichý's problem does not affect the formal structure of Stalnaker-Lewis-Kratzer theory, but rather draws attention to our lack of intuitive grip on the core notions of similarity or, for that matter, lumping. (I won't discuss here the purported triviality results Kanazawa et al. claim to have obtained in this article—they affect only some technical developments of the theory and partly for this reason Kratzer seems to have rather effortlessly managed to take the edge off them. See (Kratzer, 2005) and the footnote in (Kratzer, 1989), p.134.)

correctly predict the behaviour of such generalisations in counterfactual reasoning.

But a mere difference in lumping properties cannot bring about all the effects we have come to expect from a counterfactually relevant lawlike proposition. This is a complete list of the special properties Kratzer has finally associated with non-accidental generalisations:

1. **Weak lumping:** NAGs are true in all situations of at least some worlds.

2. **Premise sets:** Premise sets that contain only accidental propositions and no NAGs must be discarded.

3. **Confirmation:** NAGs codify information about their confirmation instances.

4. **Confirming Proposition Constraint:** Premise sets that imply confirmation instances for the NAGs they contain should be preferred.

5. **CPC for Base Sets:** The last point holds also for conversational backgrounds.

6. **Hierarchy:** Some NAGs are more natural than their equivalent reformulations.

The second point ensures that breaking a law for the sake of consistency of a premise set is a non-option.[45] The point about confirmation being hardwired in the semantic import of non-accidental generalisations is intended to account for our intuitions about Hempel's paradox and Wason's selection task. The last two constraints constitute Kratzer's attempt to explain Goodman's puzzle and Tichý's problem in terms of the privileged position that instances confirming a law play in our mind's life. This solution is chosen explicitly to avoid the recourse to causal direction—Kratzer attempts to show that 'directionality' is a feature shared by non-causal conditionals as well, and stems from the difficulties humans come up against in making inferences about non-confirmatory scenarios for a given generalisation. Here is Kratzer's King Ludwig of Bavaria example that is devised to prove this point:

> King Ludwig of Bavaria likes to spend his weekends at Leoni Castle. Whenever the Royal Bavarian flag is up, and the lights in the castle are

[45] Compare the debate between Chisholm and Rescher on (43) and (44), p.51.

> on, the King is home. As a matter of fact, the lights are on right now, the flag is down, and the King is away. Suppose now counterfactually that the flag were up. Well, then the King would be home and the lights would still be on. But why wouldn't the lights be out with the King still being away?
>
> (Kratzer, 1989), p.140

I for one have mixed intuitions about this example, however. It is not at all obvious to me that the rival counterfactual is as unacceptable as the second one from Goodman's pair. The framing of the example may admittedly give us a gentle nudge towards the option Kratzer regards as clearly more salient, but the preference is much less pronounced than she would have us expect. Moreover, it turns out that the directionality of a regularity is not as amenable to equivalent reformulations of a generalisation as the above considerations might suggest. If the answer to Goodman's puzzle only had to do with confirmation sets for the law used in the counterfactual inference, then rephrasing the law should occasion a shift in our intuitions about the counterfactual. Recall Goodman's pair:

(42) If this match had been struck, it would have lighted.

(42*) If that match had been struck, it would not have been dry.

According to Kratzer, our preference for (42) is due to the fact that in making the counterfactual inference we are leaning on the non-accidental generalisation

(84) Whenever a dry match is struck in the presence of oxygen, it lights.

This generalisation is confirmed (in the relevant sense) in a scenario with oxygen and a dry match that is struck and lights, but not in a scenario with oxygen and a wet match that doesn't light upon being struck. Kratzer argues that confirmation bias drives us towards building precisely such a scenario when assessing the counterfactual hypothesis in (42).

However, this doesn't explain why rephrasing (84) to

(84*) Whenever a match is struck in the presence of oxygen and doesn't light, it is wet.

doesn't bring the expected result of improving the credentials of (42*). This is why Kratzer is forced to add the sixth qualification to the list—she explains the asymmetry by suggesting that (84*) and its kin are not as natural as their alternatives. However, we are only offered tentative hints as to the source of this 'naturalness'.

It seems to me that at this point Kratzer becomes tangled in epicycles. There are simply too many ingenious but ad-hoc tweaks necessary to keep the theory afloat, blurring the overall picture of counterfactual reasoning. Bias towards confirmatory scenarios is, even if we eschew orthodox logical moralising, a by-product of the limits on our inferential processing power. If we place it at the core of counterfactual reasoning, we might slip back to treating the latter as a practice that is fundamentally irrational, unless we explain this bias in terms of a more interesting feature of our cognitive powers.

Besides, the approach finally chosen by Kratzer can of course only with difficulty be reconciled with examples such as (55) and (83), which bring home the fact that a handy non-accidental generalisation is not always ready at hand when people engage in counterfactual reasoning. This difficulty is shared by another early proponent of premise semantics, F. Veltman.

The distinction between laws and matters of fact loomed large already in Veltman's original take on premise semantics for counterfactuals.[46] On this approach, as opposed to Stalnaker-Lewis and Kratzer, counterfactuals weren't evaluated against an objective pool of fact, but rather against the 'opinions' of a given speaker, making them eminently subjective (roughly along the lines of the standard construal of epistemic modals). Opinions come in two sorts. Veltman called the non-accidental generalisations endorsed by a speaker her 'prejudices' (in order to stress the fact that they are not necessarily reasonable) and the matter of facts also believed by the speaker, but more likely to be given up if need be, her 'assumptions'. Into these foundations Veltman plugged the standard generator of counterfactual premise sets, with the proviso that each premise set had to contain at least all the prejudices of the speaker. The truth condition (relativised to a particular speaker) was then equivalent to Kratzer's general truth condition on page 124, with all its attendant problems. Therefore a revamped theory was put forward in (Veltman, 2005), which attempted to improve on the previous results while casting the key ideas in the framework of Veltman's later update semantics.

The new theory still draws upon the difference between laws and matters of fact supported by a speaker's cognitive state; however, Veltman no longer makes such a strong point of the subjectivity of counterfactuals—their status can be rationally adjudicated by subjecting the facts and laws that sustain them to public scrutiny. The key idea behind the new ap-

[46](Veltman, 1976), p.254

proach has to do with Tichý's Jones examples.[47] Remember that Kratzer's answer to the puzzle consisted in lumping Jones's wearing the hat with the bad weather, using the Confirming Proposition Constraint for Base Sets. Veltman aimed at the same result, but he wished to obtain it in a more principled manner.

Jones's hat was to be knocked off together with the bad weather, because the latter constituted *the very reason* why Jones wore the hat in the first place. Veltman's seminal insight was that when devising counterfactual scenarios, we need to be aware of how facts actually hang together. From a technical point of view, he implemented this insight at the level of individual worlds, but the real difference with respect to rival versions of premise semantics lies elsewhere. In order to get rid of facts dependent on others, only the situations called *bases*, construed as the minimal fragments from which complete worlds could be reconstructed with the help of laws, were to be subjected to the by now familiar revision.

First, maximal antecedent-permitting subsets of all bases are pinned down. Then they are enriched with the antecedent. Subsequently, complete worlds are reconstructed from these new situations in compliance with the accepted laws (again, more than one such reconstruction may be possible). The speaker's cognitive state will be deemed to support the would-counterfactual if the consequent winds up true in all of the latter worlds.

Veltman avoids the problems besetting naïve premise semantics by imposing the required degree of granularity on the revision process through atomic sentences. Neither does he fall in any of the other pitfalls with which Kratzer has had to struggle. However, in addition relying heavily on explicit laws, Veltman has self-acknowledged difficulties with Goodman's puzzle. While he manages to cast doubt upon Kratzer's treatment of the King Ludwig example[48], let us recall that the latter was introduced by Kratzer as a controversial attempt to show that directionality was not exclusive to causal scenarios. While you can plausibly deny that the intuitions about King Ludwig are clear enough, it is much more difficult to say the same about the original Goodman's example.

5.2.6. Conclusion

Premise semantics blew a new life into the cotenability-based approach to counterfactuals. By recasting the problem in the possible worlds frame-

[47] See page 100.
[48] See page 129

work, it has been able to answer the pesky question about the nature of the propositions that were to make up the premise sets. The new scholars have produced in-depth analyses of the notion of cotenability and have been able to shed light on counterfactual reasoning by linking it to other linguistic phenomena.

However, far too many issues still need to be addressed. Goodman's puzzle has proved particularly impervious to attempts at a satisfactory explanation. The special relationship between laws and counterfactuals has yet to be more thoroughly looked into. And last but not least, many elements of the extant theories remain in need of an independent motivation.

5.3. Causal networks

The causal approach to counterfactuals characterises a wide array of diverse semantic theories. While they can use very different formal scaffoldings, their common denominator is the assumption that at least a broad group of counterfactuals cannot be successfully analysed unless we posit that the propositions playing a role in counterfactuals reasoning are somehow structured in directed causal graphs that guide the process of such revision. As a result, especially the more formally-minded causal theories tend to outsource questions about the nature of causality to other branches of philosophy (i.e. philosophy of science or mind) and reap only the fruits in order to plug them into their semantic machinery. A passage by Pavel Tichý illustrates the spirit of this approach:

> [...] causation is a primitive, contingent relation, one which is not only not excogitable *a priori*, but also not reducible to any other, more basic notions. We shall assume that there simply exist brute causal facts, such as that event *c* is possessed of the causal power to bring another event *e*; and that to state this fact is by no means tantamount to saying that any event like *c* is invariably followed by an event like *e*, or that if *c* occurred [...], *e* would [...] occur.
>
> (Tichý, 1978), p.434

Ironically, this attitude is the source of both the enthusiasm and frustration that causal theories have engendered. The logical strength of fine-tuned causal networks enables them to effortlessly avoid the most resilient problems in the field, yielding the right results for, among others, Goodman's and Tichý's puzzle. However, just because of the power of the

tools of which these theories avail themselves, many fret that they fail to shoulder much explanatory burden. Most rival theories try to explain the behaviour of counterfactuals by showing how it arises from a few *prima facie* inconspicuous constraints—that is, for some, the whole point of semantic theorising. The worry is that most causal theories only manage to dump this responsibility on others.

I wish to argue that even where recourse to causal networks is not accompanied by further reflection on causality, there are real benefits to be obtained from exploring this possibility. To begin with, if we, as seems reasonable, use information about causal relationships in our daily predictive and pragmatic reasoning, counterfactuals may be plausibly construed as the very conventional devices developed for the purpose of conveying such information. The suspicion of irrationality, all too often raised by rival semantic theories, simply disappears. The possibility of connecting counterfactual reasoning with more mundane human activities lies now wide open.

Moreover, from the point of view of logic, not only does the causal approach allow us to formulate a logic that roughly agrees with our intuitions, but it also explains *why* we use these rules of inference and not others. This has to do with the first point—once we understand the reason for engaging in counterfactual reasoning in the first place, we can easily explain the shape this reasoning actually takes. Let us therefore have a look at how causal networks formally handle counterfactual inference.

5.3.1. Formal implementation

Any formal theory of counterfactuals that makes use of causal networks has to accomplish three basic tasks:

1. Represent existent causal relations in a formal structure.

2. Give an adequate account of counterfactual revision within this structure.

3. Show how the consequent's truth is ascertained in the revised structure.

We'll discuss each of this points in turn.

1. Causal representations

First of all, in order to apply counterfactual revision, causal links between actual facts must be adequately brought out. We have seen that mere general laws are ill-equipped for this purpose—a law such as Goodman's

(84) Whenever a dry match is struck in the presence of oxygen, it lights.

only expresses the incompatibility between four propositions about a given match (let us call them D, S, O, L). If in our match-scenario D, $\neg S$, O, $\neg L$ are true and we entertain the counterfactual hypothesis S, there are exactly three conservative ways of restoring consistency with the covering law. However, we must ensure that our analysis yields only the scenario with all the four propositions true as a viable candidate. In other words, we must implement asymmetry of dependence into our covering law—lighting depends, whereas dryness and oxygen don't, on striking the match.

There is no limit to the possible ways of achieving this result formally, so many causal theorists satisfy themselves with an informal presentation and draw on our intuitive grasp of causal laws.[49] On the other hand, on Tichý's quite formal theory[50], causality is explicitly codified as a relation between 'unfixed propositional clusters', that is, between conjunctions of (possibly temporally shifted) basic propositions that do not fix a specific time point. This relation is expressible by special connective propositions that cannot stand in causal relations themselves, but of course play a central role in determining the truth values of counterfactuals and can even feature in antecedents (most rival theories have been reluctant to go to such lengths, as the overwhelming majority of subjunctives we utter are composed of non-connective propositions).

Other theorists have leant towards a model-theoretical approach, perhaps because it strikes one as more perspicuous than Tichý's. Among the proposals in this category stand out the generalised causal models, in which each of the endogenous variables is associated with a function that determines its value based on the values of its parental variables in the causal hook-up of the system. On a practical level, given an assignment to the exogenous variables, the value of the endogenous variables is worked out after solving the system of equations that define the relevant functions.[51]

[49] See, for instance, (Parry, 1957), (Jackson, 1977) and (Sanford, 1989). Both Parry and Jackson set much store by the temporal asymmetries supposedly implied by causal laws, but this is expressly rejected by Sanford, for whom causal dependence is just one of many different kinds of dependence present in the world (temporal dependencies being one of them).

[50] (Tichý, 1978)

[51] This is the approach developed and explored, among others, in (Galles and Pearl, 1998) and (Pearl, 2000). R. Briggs follows a similar, if a slightly modified method in

A model-theoretical module can also form part of broader theories. K. Schulz has made use of Pearl's generalised causal models within the revision framework championed by F. Veltman, in order to remedy its failure to account for Goodman's puzzle.[52] S. Kaufmann implements a similar causal representation in a generalised premise semantics framework (where the conversational background splits into a modal base and an ordering source).[53]

2. Counterfactual revision

Once a representation is in place which brings out the causal dependencies between the facts pertaining to the situation under scrutiny, the next step consists in erasing everything that cannot be taken for granted in this picture in view of the counterfactual hypothesis. Note that this is nothing less than the problem of relevant conditions, that led to so much vexation among cotenability theorists and classical premise semanticists.

With causal theories, the answer to the pesky question of revision is actually fairly anticlimactic. Indeed, the purpose of the prior step of laying out the causal structure of the envisaged scenario resides in rendering the revision as trivial as this: from the initial stock, all propositions that causally depend (directly or indirectly) on the counterfactual hypothesis being false must be eliminated. Incidentally, one of the virtues of this account is a plausible explanation of the fact that subjunctives are ordinarily used as counterfactuals—there is not much point to revising with a true hypothesis.

Revision along these lines can of course be formally achieved in a number of ways. Parry and Jackson relied on *temporal* revision, that is, they assumed that by rolling the developments back to the antecedent time and throwing out the negation of the antecedent, the revision would take care of itself. But whoever believes, as one very well might, that temporal asymmetry is neither a sufficient nor a necessary condition for the presence of a causal link, will find this solution unsatisfactory.[54]

Again, this is the reason why most causal theorists feel compelled to deploy complete causal networks in the first place. Once you have described the network, the revision is conceptually straightforward. In generalised causal models, the desired result is achieved by simply generating a submodel where the relevant variable is set to the value specified by the

(Briggs, 2012).

[52](Schulz, 2011)

[53](Kaufmann, 2013)

[54]This was one of the points of Goodman's reply to Parry in (Goodman, 1957).

antecedent and its links with its parents in the model are severed.[55] This move ensures that the values of the causal progeny of the relevant variable are in turn reset. Schulz, Briggs and Kaufmann followed, each one in their own way, analogous strategies. On the other hand, Tichý didn't content himself with stopping at the negation of the antecedent and chose to eliminate also the values of its ancestry and the progeny of its ancestry. However, this may well be a case of overkill, as this account takes the side of the *backtracking* resolution of counterfactual dependence, which is non-standard by almost anyone's lights.[56]

3. Value of the consequent

Once the revision is finished, causal dependencies are again brought into play in order to reconstruct the counterfactual scenario. The new value of the variable corresponding to the consequent will determine the truth value of the whole counterfactual. In case it remains undetermined, only a might-counterfactual can be declared true.

Here, differences between the approaches we have reviewed come down to mere technicalities, as there are no conceptual grounds to prefer one over the other. Tichý was compelled to devise the notion of a connective closure (as opposed to a mere conjunction) of the antecedent together with the cotenable facts in order not to fall prey to Goodman's problem[57], but on other accounts the third step is very straightforward. For generalised causal models, it amounts to solving the system of equations that describes the submodel yielded by the revision. Schulz accomplished the same result by defining an operator on situations, each application of which determines the value of the variables whose all parents are fixed by the input. Kaufmann, in turn, resorts to an ordering source adequately reflecting causal dependencies for the purpose of enriching the revised modal base.

5.3.2. Causal logics

The counterfactual logics associated with causal networks bear more resemblance to that of the Lewis-Stalnaker conditional than to the logic of

[55](Galles and Pearl, 1998), p.155

[56]See (Lewis, 1979) for the classical discussion. An exception to the rule is Sanford, who sees nothing amiss with the set of counterfactual consequences of an assumption being inconsistent, as long as they exploit different patterns of dependence. See (Sanford, 1989), p.200.

[57](Tichý, 1978), p.442

any stricter or even material conditional. The fragment of the logic yielded by generalised causal models where antecedents are restricted to conjunctions of literals has been shown to be for all intents and purposes equivalent to Lewis's logic C1 (that is, Stalnaker's C2 without CEM but with SC).[58] Alternative approaches agree on the logic up to this point.

In particular, AC winds up invalid on all causal theories as expected. However, in addition to acknowledging the brute fact of AC's lack of intuitive support and perhaps appealing to the abstract possibility of $A \wedge B$ being somehow 'further away' than mere A, we can now state precisely when and why the truth values of $A \boxrightarrow C$ and $(A \wedge B) \boxrightarrow C$ can differ: it occurs only when $\neg B$ is part of the 'situational background' left untouched by the revision with the counterfactual hypothesis A. The fact that the background, together with the antecedent, plays a vital role in determining the value of the consequent in the new model is the very reason why counterfactuals, on their causal construal, are variably strict conditionals.

This might suggest a neat theoretical possibility of merging Stalnaker-Lewis and causal approaches by casting the similarity relation in terms of the minimal causal revision. Such a result would entitle us to the best of both worlds—the beautiful and well-behaved logic of the similarity relation and the sound theoretical grip on the meaning of counterfactuals provided by causal analysis. Nonetheless, things turn out to be more complicated once arbitrary Boolean compounds are allowed into the antecedents.

Apart from its intuitive appeal, SDA is an attractive choice for causal theorists also for independent reasons. There are three possible causal revisions associated with every disjunction (first disjunct, second disjunct and both of them) and there is no *prima facie* reason to exclude any one of them from our considerations.[59] The price, of course, is REa).[60] We are now in position, though, to appreciate why revision with A can, on these assumptions, be very different from that with the classically equivalent $(A \wedge B) \vee (A \wedge \neg B)$—there is only one minimal revision associated with the former, while there are two more complex ones associated with the latter. Thus 'logically equivalent' might surprisingly not mean 'countefactually equivalent'.

On the other hand, this paradoxically also suggests a way back to Lewis's logic without sacrificing the requirement that causal networks

[58](Galles and Pearl, 1998), p.169 and (Briggs, 2012), p.12.

[59]See (Briggs, 2012), p.18. Schulz makes SDA valid by fiat, by forcing any antecedent into a disjunctive normal form and stipulating that each disjunct must causally imply the consequent for the counterfactual to come out true.

[60]See discussion on page 76.

should codify all the information relevant to evaluating counterfactuals. We could posit that only those minimal revisions which are also the most conservative with respect to the situational background should be taken into account. That would make the revision with A identical to revision with $(A \wedge B) \vee (A \wedge \neg B)$, while keeping SDA in place for 'genuine', irreducible disjunctions (where there is no one most conservative revision).[61] The question still remains, of course, if this would be enough to accommodate linguistic intuitions. Without relinquishing SDA, however, there can be no settlement between causality and similarity.[62]

5.3.3. Philosophical import of causal theories

There must be something unsettling about formal semantics explicitly based on causal networks—the very thing that made authors such as Goodman, Kratzer or Veltman so suspicious about direct reference to causality that they were prepared to go to great lengths or even to let their theories founder rather than embrace it. And it can't only be the fact that causality is itself in need of philosophical clarification—most of the authors wary of it have employed concepts (such as that of natural law, similarity of possible worlds or lumping) that are not much more transparent than that of causal dependency. Where does the squeamishness about causality hail from, then?

I suspect the answer has to do with the limits of formal semantics. If causality is indeed the basic, irreducible linchpin of counterfactual reasoning Tichý would have it be, then it seems to me that the formal tradition in semantics is bound to forfeit its central role in elucidating the meaning of subjunctives.

We can get a clearer view of the problem if we first try to answer the question as to why (mostly first-order) set theory or lambda-calculus should be expected to provide the universal language for elucidating linguistic meaning. Put like this, it surely does not sound as self-evident as it might given the overwhelming tacit consensus in the field. Yet there is a good reason—both set theory and lambda calculus are mathematical tools perfectly suited to creating complex structures out of the meagrest of

[61] This is admittedly very impressionistic and still unsatisfactory as an attempt at a convergence of the two logics. While this suggestion would causally validate $(A \vee B) \boxrightarrow C \vdash (A \boxrightarrow C) \wedge (B \boxrightarrow C)$ for atomic A, B, on Lewis's semantics this would fail whenever we had a D such that $A \boxrightarrow D$ and $(A \wedge B) \boxrightarrow \neg D$ come out true in the model.

[62] Some even argue there is no principled way of achieving this and that closeness of possible worlds, despite all its vagueness, is an inappropriate notion for the analysis of counterfactuals anyway. See (Fine, 2012), p.38.

resources. They are therefore well-equipped to spell out the contribution of a sophisticated language device in terms of its systematic effect on several more easily fathomable variables. That's how formal semantics can be useful in elucidating meaning.

However, I suspect it would do a lot of good to the current mainstream in philosophical logic and formal semantics if the limits of this approach were more systematically explored. A semanticist is often tempted to use variables whose workings are by no means as transparent as they would have to be for the analysis to be in any way enlightening—we have seen quite a few examples ourselves. Moreover, it is highly unlikely that all our interesting vocabulary will be susceptible of analysis in terms of a few unproblematic magnitudes.

In the case of causal theories of counterfactuals, the actual formal implementation, virtuosic though it may be, doesn't provide much new semantic insight. No exciting new theorems are in the offing and any theoretical surprises will most likely cast doubt on the aptness of a particular implementation rather than give pause to our intuitions about counterfactuals.

The issue of backtracking is a case in point. We saw in Tichý, and especially in Kaufmann, that whether or not the system will support backtracking is a conscious modelling choice that is made by a theorist in accordance with his prior views on the subject. We can no longer expect the formal system to give us an enlightening answer to the problem—the presence or not of backtracking will not in general be a consequence of any more basic assumptions. With causal accounts, all important decisions about the notion of causality must be taken before the very construction of the formal system—this time round it is the chief pillar and not the whole structure which commands most of our interest. However, the investigation of an ingredient in a formal analysis must perforce take a different shape from that of the entire architectonic complex.

Chapter 7 is supposed to provide a first stab at such an analysis. Once we have acknowledged the edge the causal account enjoys over its rivals, I submit that it behoves us to take a closer look at how causal patterns are put into play in real life scenarios, how they relate to indicative conditionals and other linguistic expressions, and to check whether they are capable of explaining away at least some of the puzzles (other than Goodman's one) we have encountered along the way. What is more, while causality does play an overwhelmingly important role in counterfactual reasoning, some counterfactuals do not easily fit this mould and some are plainly not causal.[63] If we really aim to get a sound grip on counterfactuals, we have

[63] As an example of the former, think for instance of the dill sauce example in (83). With

to face up to these outstanding issues.

In its entirety, of course, this would be a gigantic undertaking which I am not in a position to carry out. However, I would like at least to defend the viability of such a project by showing some early theoretical fruit it can bear and by paving the way for other scholars tempted to venture away from the formal orthodoxy.

regard to the latter, as even D. Lewis had to admit (see (Lewis, 1973a), p.24), mathematical reductio is often set forth in counterfactual terms.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Part II

Conditionals in natural language

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 6

Conditional speech acts

We have seen a plethora of very heterogeneous theories of conditionals. It is high time to take stock and try to answer at least some of the many questions we have encountered along the way. I think the best way to do it is to start from the very beginning and ask ourselves what we should actually be aiming at when putting forward an 'account' or a 'theory' of conditionals.

Admittedly, the question may sound a little outlandish. This may not be because there is anything inherently awry about it, though, but rather for the simple reason that very few writers stop to give it a minute's thought before getting on with their own 'proposals'. We are expected to have a tacit understanding of the nature and goals of the undertaken endeavour, so it seems redundant to state them in full.

But observe that terms such as 'account' or 'approach' are, on their own, far from self-explanatory. What would you expect from 'a theory of' or (even worse) 'an approach to', say, the motor vehicle or tropical rainforests? There are myriads of angles from which you can inspect either of these phenomena. For instance, you might focus on their physical structure, dynamics, history, or wider social, economic or environmental implications. Each of these aspects in turn requires a particular theoretical approach, a choice of relevant variables and an adequate level of abstraction. That's why you don't tend to encounter 'theories of X', unless it is quite clear what the ultimate aim of such a theory would be in a given context.

Is this the case with philosophical theories of conditionals? I have my doubts. Just as cars or jungles, language phenomena can be analysed in a number of different connections. I shall argue that it is by no means clear which ones are actually intended by most theoreticians, and that this in turn leads to quite some philosophical confusion. By throwing some light

upon this question, I hope to show that the current impasse in philosophical semantics may be due to an inadequate methodological attitude.

6.1. Meaning of conditionals

You may have found the preceding remarks disingenuous. It is surely no secret what most theories of conditionals are after? The default assumption in the literature seems to imply that the ultimate goal of our theorising is a correct rendering of the *meaning* of conditionals. However, there are reasons to think that it may as well be one of those cases when people fail to characterise a task that they otherwise carry out proficiently.

That there *is* something amiss with our understanding of the aims of the philosophical inquiry into conditionals is witnessed by the recurrence of one surprising claim. The proliferation of different theories and the failure to reach a reasonable theoretical consensus on quite a few resilient puzzles is sometimes interpreted as a sign of our ignorance about the meaning of conditionals.[1]

I find this to be a particularly infelicitous way of putting things, chiefly because under its literal interpretation the claim is blatantly false. Of course, in order to know what a certain expression means you do not have to be able to explain its meaning by means of a non-circular paraphrase—it will suffice if you can successfully use it in communication. Most of us would be in a quandary if asked to explain the exact meaning of 'cynical', but our skill in applying this label to different human actions is all that is needed to know what it stands for. Knowledge of meaning is thus basically a kind of knowing-how rather than of knowing-that.

This also applies to conditionals—there can be no doubt that competent speakers do know what they mean, even if they are not ever in a position to convey this knowledge explicitly. The situation is, thus, not as bleak as it might look at first sight, and the scaremongers responsible for the confusion are of course perfectly aware of this. What they really seem to be up to is to draw the attention to our lack of explicit theoretical understanding of the meaning of conditionals. And in order to achieve this, they choose a striking, if slightly misleading, formulation.

The failure to provide such an account needn't be a big deal from a practical point of view. We can live without a conclusive specification of what makes us label an action cynical or an utterance ironic. And as far as I now,

[1]For explicit avowals, see for instance (Mackie, 1973), p.64; (Kratzer, 2012), p.85; or (Bennett, 2003), p.2

no philosopher has been involved in the project of clarifying the meaning of 'tawdry' or 'underpants' (although in view of the current splurge in public funding for philosophical research, I may easily be wrong). And sometimes there is simply nothing of relevance to be said about the meaning of a particular expression—though apples themselves may constitute quite an interesting object of study, 'apple' means what it means and that's about it. And sometimes what little there is to say about such a meaning is clearly a job better suited for a linguist rather than for a philosopher.

Why should the case of conditionals be any different from these examples? What makes their meaning deserve a properly philosophical analysis? A part of the answer is historical—questions concerning meaning of conditionals were present at the outset of Frege's new logic and as such have secured a pride of place within the analytical tradition.[2] Yet more importantly, it is often felt that the use of conditionals is linked to some of our most fundamental cognitive processes. That's why it is surmised that their study might lead to new profound insights into human reason.

Let us provisionally grant, for the sake of the argument, that the mission of such study is to elucidate the *meaning* of conditional sentences. We can then ask what kind of entity meaning is and how we usually go about the task of explaining it. Such a reflection will help us understand why so many of the theories we have inspected show *ad hoc* features.

6.1.1. Truth-conditional semantics

Semantic reductionism

There is a default answer to the question about the proper way of proceeding in semantics. According to the received view, the correct form for presenting the meaning of an expression is by pinning down its contribution to the truth conditions of typical utterances of sentences in which it appears. And if the expression, as in the case of a conditional, is a sentence-forming operator, we can simply speak about its own truth conditions. The idea is, of course, that the meaning of an expression defines (and is defined by) a function from the semantic values of the rest of the sentential ingredients into truth values. Formally

Truth conditions of a binary sentential operator
$[\![\Phi(A, C)]\!] = f_\Phi([\![A]\!], [\![C]\!])$, where $f_\Phi : \mathcal{P}(K) \times \mathcal{P}(K) \to \{0, 1\}$

[2] Although, as we have seen in Section 1.2.1, this tradition largely distorted Frege's and Russell's original thoughts on the subject.

But is this really the best way of explaining linguistic meaning? Remember that when translated into a little less mathematical language, the above truth condition takes a form resembling that of a Tarski's T-sentence:

Truth conditions of a binary sentential operator
$\Phi(A, C)$ is true iff p.

Here, p is a sentence of the metalanguage identical in meaning with the sentence $\Phi(A, C)$ of the object language (it can, and probably will, contain the metalanguage analogues of A and C as parts). In Tarski's case, the former was a direct translation of the latter. This was all that was needed for the purposes of Tarski's project, which *did not* aim to analyse the meaning of the left-hand sides of T-Convention instances, but rather to provide an adequate formal definition of the truth predicate.

If you however try to get something like a T-sentence to display the meaning of a natural language sentence, you had better be aware of a few difficulties. First of all, your metalanguage will in general be identical to your object language. There is no *prima facie* reason to privilege any subset of our vocabulary as particularly suited for the task of explaining meaning; in any case, this would require an independent argument. In the absence of a clear distinction between object and metalanguage, the statement of a truth condition amounts to little more than a mere paraphrase.

Another related point is that for the analysis to work, you must avoid trivial paraphrases. A direct translation won't do any more as the fact that it is used as opposed to mentioned doesn't confer any added value on the analysis. The proposed equivalence should somehow be informative. Yet it is not at all easy to pin down what this informativeness is supposed to consist in. Natural contexts for explanations of meaning all include an interlocutor who does not understand the expression at issue. Then it is obvious that the analysans must conform to his or her limited linguistic knowledge—an explanation that works in one context might fail to do so in another. But how to concoct an interesting and informative semantic analysis for the general public, conversant with the meaning of the analysandum, is by no means clear.[3]

[3]Heim and Kratzer try to assuage these worries by pointing out that formal semantics doesn't really purport to explain the meaning of our vocabulary, but only model its compositional properties: (Heim and Kratzer, 1998), p.2. However, as regards conditionals, this seems to contradict not only the widely shared desiderata in the field, but also Kratzer's own complaint to the effect that the absence of a successful account implies lack of grip on the meaning of conditionals. The purpose of the theoretical part of this dissertation was to show that the triviality of Tarski's T-schema didn't necessarily vanish on the level of sentence-forming operators—it was often only disguised by the complexity of the analysis.

And last but not least, there is absolutely no guarantee that there *exists* an adequate non-trivial paraphrase in the language for the piece of vocabulary you are interested in. Actually, it seems that the more interesting expression you choose, the less likely you are to find one.[4] In particular, most scholars try to avoid the problem of triviality by constraining the analysans to a proper subset of vocabulary deemed conceptually more basic than the analysed expression—but there is no *a priori* reason to expect that the latter is thus reducible.

Why have we been so keen to accept truth-conditional approaches to meaning, then? The answer resides in the enduring allure of what I have taken to calling 'the Switzerland conception of language'—it transports us into a magical land, where all activity is subject to strict rules and even the wilderness is organised in neatly designated and strictly regulated areas. This conception assumes that you can oversee and control the whole process of belief creation, along with the way how beliefs are translated into language. Such a conception, for a part, arises naturally (we like order anyway, because it gives us grip on things, and indeed, many linguistic neighbourhoods *are* well-behaved); however, it has been given a great boost by the rise of logical positivism. Its champions claimed that we had access to utterly unquestionable facts (which ironically were conveyed by the most subjective of experiences) and all our meaningful utterances directly or indirectly made assertions about them. In other words, they posited a linguistic level (that of direct observation reports) in which language directly connected with the world and was thus the source of all meaning. All higher linguistic levels were constructed logically from the basic one and derived all their meaning from it.[5]

Logical positivism is no more in vogue nowadays, but I think many of our theoretical tics can be traced back to this source. Semantic reductivism seems to be inspired by the very idea that your analysis can eventually reach an unproblematic linguistic level, where the truth or falsity of your

[4]This is the point that T. Williamson very convincingly makes against the traditional epistemological programme of truth-conditional analysis of the verb 'to know' in (Williamson, 2000), p.31. It seems to clash, though, with what D. Blakemore has to say in (Blakemore, 2002), pp.82-83—I don't think we should be able to provide paraphrases for all conceptual words in our language, nor that paraphrases for procedural words are *a priori* impossible. I do agree, though, as will presently become clear, that a description of the actual use of a word goes in general a longer way towards explaining its meaning than a paraphrase.

[5]This is meant as a rough general description that abstracts from the particularities of the philosophical systems comprising the movement. For an excellent exposition of the epistemological and semantic ideas that constitute the backbone of logical positivism, see (Williams, 2001), chapters 7 and 8, and especially p.91.

assertion can be read directly off the world. So a truth-conditional analysis of an expression can lead you back to this linguistic eden, where life is tranquil and devoid of worries. Language, on this construal, is essentially and unproblematically descriptive.

We shall see plenty of evidence in the following chapter suggesting that even what we regard as simple descriptions of the world are often defeasible and theory-infused mental constructions, only partially underwritten by sensory experience. In the meantime, though, we can inspect some of the unpalatable results semantic Switzerlandism has brought in its wake.

Just think of the mind-brain identity theory (not to be confused with the identity theory about indicative and material conditionals), so popular in the 1950s. I submit that nobody would ever have found it a good idea to identify the talk about emotions, personality traits and intentions with the talk about particular brain states, if we did not subscribe to a conception of language according to which all statements that are not directly evaluable as to their truth must be reducible to such statements.

Or take the meaning of normative vocabulary, such as that of 'interesting'. If in your semantic analysis you only are permitted to avail yourself of unproblematically descriptive vocabulary, the meaning of 'interesting' all of a sudden appears shrouded in mystery. A request to fill in the gap in '*X* is interesting iff ...' prompts one to look for features that make things deserving of our attention—and given the range of *X*, this will necessarily be a next-to-impossible endeavour, which can all too easily lead to semantic nihilism. That's why logical positivists took the other horn of the dilemma, declaring normative vocabulary to be merely expressive of speaker's attitudes and emotions and as such unfit for serious debate about truth and falsity. But this is unnecessarily restrictive—a better solution would be to make a sharp distinction between theoretical accounts of the meaning of a sentence such as 'X is interesting' and theoretical accounts of when the utterance of such a sentence is warranted. The point is that naïve descriptivism conflates these two endeavours.

Simplistic descriptive conceptions of language were subjected to searing criticism throughout the 20th century. Nowadays, the study of arguments advanced by Wittgenstein, Quine and Searle is part and parcel of any analytic philosopher's training. However, model-theoretical methodology, as applied to natural language, has been very successful in inadvertently promoting return tickets to Swiss Alps. This is because regardless of the nature and complexity of your models, the methodology guarantees that there *is* vocabulary for directly describing them. You only have to provide translation rules (truth conditions) for the expressions you have set out to study, and will have succeeded in analysing their meaning.

However, what we said above implies that the application of this methodology to natural language should be viewed with a great deal of suspicion. The concept of 'natural language semantics' is much less innocent than it might appear at first sight.

Truth conditions of conditionals

The foregoing can help us put Lewis's and Hájek's triviality results and Edgington's arguments against truth conditionality of indicatives in proper perspective. The bottom line is not that we have to deny indicatives truth conditions if we want to preserve the intuition behind the Equation—for any expression whatsoever one can of course put forward trivial truth conditions:

Indicatives trivial truth conditions
$A \rightarrow C$ is true iff C if A,

or some more sophisticated variant featuring, say, 'on condition', 'provided that' or 'should' with inversion.[6] Note, though, that the triviality of these more involved paraphrases does not ensue automatically, that is, you couldn't for instance read it off the syntax of the analysans. It is only because we regard it as uninteresting that the analysis 'fails'. This suggests that what we are really (though perhaps only unconsciously) after in this kind of semantics is some kind of reduction after all.

The correct conclusion to be drawn from triviality proofs, it appears to me, is that of Egré and Cozic.[7] They interpret Lewis's and Hájek's proofs as inexpressibility results. In other words, according to Egré and Cozic, these proofs show that a non-trivial language containing conditional probability operators has strictly more expressive power than its fragment that only contains unary probability operators. Thus if conditional probability really plays a crucial role in the assertibility of indicatives, they are not in general reducible to common non-conditional propositions. So at least a wide range of possible paraphrases turns out to be inadequate. On the other hand, we have seen that many of the adequate paraphrases are too close to the meaning of indicatives to be any useful for our purposes. This should made one at the very least wary about the possibilities of a successful truth conditional analysis of indicatives.

[6] Pace (Edgington, 2007), p.162.

[7] (Egré and Cozic, 2011). A similar interpretation is also suggested in (Milne, 1997), p.196. See Section 2.3.1.

Is the situation that different with subjunctives? Most scholars have answered affirmatively. According to this way of seeing things, while indicatives lack truth conditions because they are basically devices for expressing subjective conditional probabilities, subjunctives are regular propositions through and through and definite truth conditions can be stated for them. In what follows, I will argue that this distinction, though based on a few sound intuitions, sows more confusion than it provides insight.

Of course, just as with indicatives, you can also formulate trivial truth conditions for subjunctives.

Subjunctives trivial truth conditions
$A \mathrel{\square\!\!\rightarrow} C$ is true iff C would have been the case, if A had been the case.

And, as before, we can find almost-trivial reformulations, formed for instance by inversion, 'but for', as well as with 'provided' or 'on condition that'. These sound even more superficial than their indicative analogues, though, as most of them must use the same grammatical tenses that are peculiar to the core cases of subjunctives.

How about non-trivial truth conditions for subjunctives? There are in any case no triviality results that would *a priori* preclude such a possibility. And such a project surely must seem viable in view of the extensive literature on the subject. Or does it?

One of the aims of the preceding chapters was to make the reader aware of how badly the most celebrated theories of subjunctives fare on that score. Once we manage to filter out all the noise about logic, nature of possible worlds and all kinds of syntactic quirks, we usually end up with a semantic analysis that makes a heavy use of some suspiciously intangible variable that is doing most of the dirty work. Cotenability relations, similarity orderings, lumping properties, envisaged contexts, and (admittedly to a much lesser degree) causal networks—these are often little more than clever ways of obscuring the fact that we are essentially back to the trivial truth condition stated above.[8]

Of course, our lack of success in the quest for interesting truth conditions for subjunctives by no means implies that such an endeavour is doomed to failure from the start. However, I think that it at least should make us question the default assumption that such a reduction should be possible or even *desirable*.

But why do most theorists then make such a stringent distinction between the allegedly non-propositional indicatives and the apparently

[8] I'll have more to say about causality in what follows.

much more conventional subjunctives? The answer, that underlies all triviality proofs, is that the belief in $A \rightarrow C$ cannot be severed from the belief in A and C—uncertainty about the former does not go beyond uncertainty about the latter, as is the case with, say, 'because'. Fully determined states of affairs, like those that we use in semantics for our logics, therefore end up trivialising indicatives. A subjunctive '$A \boxdot\!\!\rightarrow C$', on the other hand, is largely independent of whether A or C is the case (except when $A \wedge \neg C$ happens to be true), so it usually can be added to the welter of propositions supported by a state description without causing unwelcome interferences. This means that counterfactuals can be made to work within our semantic framework in a conventional way.

This is all very well. However, we should guard ourselves from overstating the explanatory power of the classical possible world semantics—it is useful for modelling certain linguistic phenomena mainly because of the many simplifications it is subject to. It surely should not be construed as having the last word in inquiries into meaning. In particular, it does not imply much about the question whether a successful statement of non-circular truth conditions is possible in practice. It can be of use in proving the negative result, as we have seen, but it is not sufficient to guarantee a positive one. To be sure, if we had linguistic resources to single out each and every possible world, we could use a mere disjunction—but atomic possibilities are just a convenient fiction without clear real-life counterparts. And even if we had such resources, nothing guarantees that they wouldn't include subjunctives or their trivial variants.

Be it as it may, in view of the rather disappointing results of the traditional analyses of conditionals, I think it is high time to approach the study of their meaning from a different angle. In order to do so, we first have to become clear about what we actually understand by meaning and what the standard procedures for explaining it are.

6.1.2. Meaning and truth

In traditional semantics, the meaning of a sentence is equated with the function that allows us to determine its truth value in any given state of affairs.[9] The meaning of a subsentential item is then defined as the function obtained from the original one by λ-abstraction on the variables corresponding to the other items that have to complement it in order to yield a propositionally evaluable content.

We have already seen one problem with this strategy—when trying to

[9] Witness the offhand first sentence in (Heim and Kratzer, 1998).

specify the function that constitutes the meaning of an expression, it is all too tempting to smuggle in vocabulary trivially synonymous with the analysed item. This objection, though, does not get to the core of the above way of looking at meaning—in endorsing the traditional way of doing semantics, you don't have to posit that we always have an interesting way of spelling out the function in question (as a reductionist would require).

A more direct way of attacking this conception of meaning is to argue that it accords a disproportionally central role to the notion of *truth*. This, of course, is not to say that truth is irrelevant in semantics—far from it. Discrepancy in truth value of two different sentences in certain circumstances of evaluation is an infallible sign of a divergence in meaning; however, not all differences in meaning can be isolated in this way. This is the case with Grice's conventional implicature: you can replace an 'and' in a compound sentence with a 'but' and the result is guaranteed to share the truth value with the original sentence. However, this will of course be at the cost of some awkwardness due to the different meanings of the two conjunctions. The upshot is that mere truth is not fine-grained enough to reflect all the semantic shades our vocabulary can and does adopt. Putting forward truths may be a central, or even the ultimate function of natural language, but it is by no means the only one.

Escaping from truth as the main tool of an inquiry into meaning requires broadening our view of language. The conception of meaning as a function that takes sentences and circumstances of evaluation into truth values completely abstracts from people who use the language and their reasons for doing so. This comes extremely handy in logical calculi, which are intentionally divested of non-truth-conditional elements, but is too schematic to suffice for a serious study of natural language. In order to explain the meaning of a whole host of expressions, one must take into account at least some of the quirks of human communication.[10]

Take 'but', for instance. Both for Grice and Bach, an utterance of 'A but B' conveys two propositions: on the one hand, $A \wedge B$, and on the other, the idea that there is some contrast between A and B.[11] For Grice, the second proposition is a conventional implicature and doesn't form part of the truth conditions of the utterance, whereas for Bach it does, despite being less focalised than Grice's minimal proposition. I find neither of the two positions very convincing, and I think the reason is that they cling too heavily to truth as the ultimate semantic explainer. If there is a difference

[10]For a brief overview of non-truth-conditional vocabulary, see (Blakemore, 2002), pp.32-38.

[11](Grice, 1975) and (Bach, 1999). A similar view is defended in (Rieber, 1997).

between 'and' and 'but', the reasoning goes, it must show at the level of truth or falsity of some proposition, whichever status it may enjoy.

But I submit that this approach to 'but' is implausible. If you say

(85) She is poor, but honest,

you are not in general trying to imply, let alone assert, that there is a contrast between the two predicates. What use could anyone have for such a vague and, frankly, dull piece of information? If your audience perceives the contrast, mere $A \wedge B$ will be enough. And if it doesn't, you are not likely to achieve much by simply saying that there is one. A much better explanation is that you use 'but B' when you somehow want to restrict the range of conclusions your audience might reasonably be expected to draw from A or an assumption made manifest by A.[12]

Note that an account along these lines makes an explicit reference to speakers, audiences and the interaction between the two. It conceives speakers as agents trying to produce certain effects in the audience, who, in order to succeed, must carefully calibrate their words. Language is an instrument that enables speakers to achieve some of these goals—and if there is a particular effect that is frequently intended, it is only to be expected that language will provide conventional means of accomplishing it.

What is more, this approach is useful even for vocabulary with truth-conditional meaning, because it urges us to assess the contribution of scrutinised expressions on its own merits, thus discouraging reductivism.

Take the example of normative vocabulary discussed in the previous section. As I have already suggested, I think that the right way out of the quagmire with normative vocabulary such as 'interesting' is to make a distinction between the meaning of 'interesting' and a theory of interesting things. This meaning is rather straightforward if you are allowed to explain it by appealing to the function 'interesting' standardly performs in a conversational exchange. Roughly, by labelling something as interesting, the speaker puts it forward as worth engaging with or intellectually stimulating. An investigation into our motives for finding something worth engaging with can be a respectable occupation for, say, a neuroscientist

[12]This is intended as a reference to the procedural theory of the meaning of 'but' by D. Blakemore in (Blakemore, 1987) and (Blakemore, 2002), pp.98-115, but there are many variants that also fit the somewhat crude formulation I have given here. Actually, traditional linguists have for a long time been aware of the range of pragmatic variables that must be taken into account when explaining linguistic meaning—see (Ducrot et al., 1980), pp.93-130.

or an anthropologist, but has little to do with semantics.[13] You may even conclude that given the wide array of human interests and their high contextual sensitivity, no interesting theory of interesting things is as much as possible. However, this would not make the meaning of 'interesting' any more cryptic or ineffable.

I think that a story along these lines can also be told about conditionals. Their meaning is one thing; how we come to believe them, when we are disposed to abandon them and how we justify them, are related, but different issues.[14] In formal languages, these are two sides of the same coin. However, inferences we legitimately draw in natural languages depend on much more than the meaning of the expressions involved. One of the central theses of this dissertation will therefore be that in natural language, meaning alone does not determine logic. And I'll also argue that there is nothing like 'the genuine logic' of conditionals, as each formal system has to make a pragmatic choice between the inferential features it sets out to model and others that must be left out. As always with abstraction, the one involved in logical modelling increases the precision of our inquiry by reducing its focus.

Meaning, on this view, is an abstract notion that points towards the standard function that a piece of vocabulary performs in a conversational exchange. Remember bias? It was also an abstract concept devised to facilitate talk about cases of unfair treatment. You could not touch it, point your finger towards it, and what you *really* did not need was an unorthodox metaphysics to warrant talking about it. I suggest a similar picture is roughly applicable to meaning as well—it is a notion designed to streamline exchanges about the correct use of vocabulary. And there is no default way to explain how a particular expression should properly be used.[15] With a speaker of another language, you may provide a direct

[13]This is the very reason why explaining attributions of mental states as making an implicit statement about brain properties, as in some crude versions of the identity theory in epistemology, fails as an account of the meaning of such attributions; this, of course, without prejudice to the obvious fact that brain states are vital for having the mental states in question.

[14]Fortunately, I am not alone. The following passage is by T. Honderich: '[...] there are *two* questions, one of them that of specifying the *meaning* of conditionals, the other the general analysis of their *grounds* or *premises*. It is in fact this latter problem to which philosophers have addressed themselves, despite their misdescription of it, and they have had arguable things to say about it. Their efforts must be seen in a proper light, not the one they supply.' (Honderich, 1982), p.300.

[15]I admit that the extensive literature on the problem of normativity (very few issues fail nowadays to spawn extensive literatures) takes me out of my depth and my ideas on the topic are probably too blunt. Anyway, by 'properly' I simply mean something like

translation (if there is one), otherwise you can try to find a paraphrase using vocabulary she does understand, or a stab at a second-order explanation of the linguistic function of the expression at issue. But gestures, especially pointing towards relevant objects, and grimaces can turn out to be just as effective.

An important point is that such explanations are not meant to be exhaustive. Just as you won't learn to swim only by listening to your coach's instructions, helpful as they may be, you are usually not expected to be able to use an expression perfectly just by understanding someone's clarification. Several rounds of trial and error might be needed before you become proficient—we infer the meaning of most expressions, certainly in our native language, without need for explicit explanations anyway. We are so good at grasping meanings of words because the conversational functions they serve strike us as natural. They are, as it were, cut out to satisfy human communicative needs.[16]

For all the above reasons, I think that a more promising way of approaching the meaning of conditionals is by taking a look at the function they are designed to carry out in real life communication. A review of such a function need not suffice to provide a semantic reduction, but this is no longer our aim anyway.

6.1.3. Conditional Speech Acts

John Austin famously reminded philosophers who still observed natural language through the prism of formal calculi that we can do a lot more with words than just assert descriptive contents. We can, for instance, also ask for information, issue orders, give advice, make pledges, threats and pronouncements, offer excuses, take oaths and undertake commitments. We say that such utterances differ in their *illocutive* force. While with mere assertions it makes sense to ask whether they are true or false, utterances with other illocutive forces cannot always be felicitously characterised in this way. Questions can at most be pertinent, advice useful, orders fulfilled and promises kept. But truth *can* be useful in analysing such acts: questions can be answered affirmatively if their propositional content is true, advice is useful if it is true that following it leads to the intended outcome, a

'effectively with respect to your audience'. I am inclined to chalk the difficulties with this idea up to the need for broader linguistic coordination within a community.

[16] That's why I have never fully understood the interest of Davidson's project. What use do we have for an abstract theory of meaning that has nothing to do with the way people learn meanings in the real world? Doesn't it necessarily lead to a skewed conception of language?

promise is kept if its content becomes, or rather is caused to be true at an explicitly or implicitly specified later point. It is therefore reasonable to assume that the standard function of a piece of vocabulary should not wary wildly across utterances with different illocutive force. I would like to make use of this intuition in the analysis of conditionals.

Indicative conditional clauses can appear in almost any kind of speech act. Let us see a few examples:

Conditional advice and orders

(86) If you see the gabby professor approaching, run for your life!

(87) If you come under artillery fire, retreat immediately.

Conditional promises

(88) If you can spare a penny for my campaign, you will very soon see some juicy tenders coming your way.

(89) I'll pay you back this very week, provided he buys the stash.

Conditional excuses and pronouncements

(90) If the late-night karaoke party at my place caused you any trouble, I am very sorry.

(91) If there are no objections by those present, I pronounce you man and wife.

And there are even speech acts that are by their nature almost always conditional, such as threats:

Conditional threats

(92) If you don't cough up the cash, something nasty will happen to your little rubber duck.

I believe we can distinguish the function the conditional clause performs in most of the above examples very clearly—it seems to constrain the scope of the speech act carried out by the utterance of the main clause. The advice to run for your life is meant to be heeded only upon seeing the talkative faculty member, the order to retreat only becomes binding in the event of enemy bombardment, and of course your company may have to wave the public contracts goodbye if you do not chip in any cash into the party chest. Was the advice followed, the order fulfilled or the promise

kept if the professor and the enemy gave you a break, and if the campaign had to do without your contribution? The question sounds strange—we feel that in such circumstances there was no advice, order or promise to begin with.

The almost ridiculously trivial suggestion I would like to make here is that the standard function of if-clauses is exactly the same in assertive utterances as in utterances with other illocutionary forces. On this view, if-clauses limit the scope of the assertion corresponding to the main clause of the sentence, so that it only enters into force once the condition specified in the antecedent is satisfied.

Of course, I would not dare to make any claim to originality with this simple theory. Actually, it was given a much neater presentation more than half a century ago:

> Now under what circumstances is a conditional true? Even to raise this question is to depart from everyday attitudes. An affirmation of the form 'if p then q' is commonly felt less as an affirmation of a conditional than as a conditional affirmation of the consequent. If, after we have made such an affirmation, the antecedent turns out true, then we consider ourselves committed to the consequent, and are ready to acknowledge error if it proves false. If on the other hand the antecedent turns out to have been false, our conditional affirmation is as if it never had been made.
>
> (Quine, 1966), p.12

We witness once again that classical authors had a much clearer idea what they were up to than many of their contemporary peers.[17]

The anticlimactic contention that I also want to advance is that Quine's intuition is a basic fact about the meaning of indicative conditionals, which is not to be explained by any deeper semantic insight. We have seen that it is not necessarily the case that the inquiry into the meaning of an expression must lead to surprising discoveries—you don't expect them in the case of 'apple', so why should you in the case of 'if'? Moreover, I think that even on such a meagre basis, we can explain at least some of the more enigmatic features of the behaviour of indicatives, such as the different logics to which they give rise and their interaction with quantifiers.[18]

[17] See (Sanford, 1989), p.63, where Quine is interpreted as adopting a position on the material conditional akin to the one we have attributed to Frege and Russell.

[18] Given the simplicity of this view and its copious theoretical advantages (to be presented in what follows), it is baffling how little following it has attracted—this attests to the powerful grip of the traditional way of doing semantics. Among the authors who

Conditional questions

But before moving on to those points, a few remarks are in order. For a start, are there conditional questions? I have been conspicuously silent on this issue, and for a reason—it seems to me that most examples of conditionals embedded within questions we can come up with correspond to questions about conditionals rather than genuine conditional questions:

Conditional questions?

(93) What will happen if I press this red button?

The question here is not subject to the antecedent obtaining—the very answer to this question may prevent you from pressing the button in the first place. You seem rather to be asking your audience to make an assertion about what will happen, conditional on the assumption that you will press the button.

We don't seem to have much use for genuine conditional questions in our language. We ask questions typically when we are in need of particular information that our interlocutor can be expected to provide; if we subject our questions to a condition, our interlocutor may simply no longer be around to answer it when the condition becomes (known to be) fulfilled. But this does not mean that they don't or even can't, strictly speaking, exist. We get somewhat closer to conditional questions with a construction such as:

(94) If he went to music school, how can his singing be this bad?

This is a peculiar (though proper) use of a conditional clause whose purpose is to signal incredulity. Compare the following assertion:

(95) If *he* can work as an official interpreter, then so can my pet parrot.

---

have explicitly defended this view, D. Edgington stands out—however, she subscribes to a rather traditional conception of truth, which detracts a little from the punch of what may otherwise be the most enlightening theory of indicative conditionals on the market: (Edgington, 2007). M. Woods has also defended it in (Woods, 1997), but he excluded predictions from the scope of his theory, classifying will-indicatives, following Dudman, together with subjunctives. DeRose and Grandy have applied it to the analysis of biscuit conditionals (see Section 6.3.2), citing Quine as precursor: (DeRose and Grandy, 1999), p.408. There is a useful historical review of analogous positions, although concerned primarily with their mathematical implications, in (Milne, 1997), pp.197-212. All in all, many theories of indicatives are not put forward as strict alternatives to the conditional-assertion thesis, but rather as its elaborations, probably because such a meagre account is deemed insufficient for the task of explaining meaning (see for instance (Jackson, 1987), p.13). This is the assumption I am trying to question in this dissertation.

Here you condition your question and your assertion on the antecedent obtaining, pretending that you still do not know for sure whether it really does. The above-mentioned problems with conditional questions are naturally circumvented, since assuming the antecedent *does* hold, your question is for all effects and purposes immediate.

A completely genuine example of a conditional question might go as follows: suppose your boss entrusts you with interviewing a sales representative from another company you do not know much about and you are to fill him in afterwards. He gives you the following note:

(96) If they sell rubber gum, what density is it? And if they sell chewing gum, what flavour is it?

Of course, you are only expected to answer either question if the corresponding antecedent is true.[19]

Subjunctives

Another point concerns subjunctive conditionals. In English, they are usually distinguished from indicatives on a syntactic level by the presence of a past modal in the main clause and a tense shift in the subordinate clause. When such a modal has an independent non-conditional use, it can of course be used indicatively, as in

(97) If he is as good as they say, he might give a boost to our team.

or

(98) If he discovered the truth about the President's dirty secrets, he would be dead in a week's time.[20]

However, in genuine subjunctive cases the main clause is not autonomous and requires at least a tacit subjunctive antecedent.[21] That's why plugging an indicative conditional clause to a subjunctive main clause doesn't usually produce a meaningful sentence:

(I-S*) If he knows how to solve this exercise, he would not flunk the test.*

The upshot is that subjunctives really make up a distinct category, and are not just 'woulds restricted by a conditional clause'.

[19] For a similar stance on conditional questions, see (Woods, 1997), pp.76-77.
[20] The example is adapted from (Bennett, 2003), pp.10-11.
[21] Compare (Woods, 1997), p.9 and (Veltman, 2005), pp.160-161.

The job of the main clause of a conditional sentence is to specify the illocutionary force of the utterance. Therefore, a genuinely subjunctive 'would' seems to rule out such speech acts as orders and direct advice (which require the imperative mood), promises (typically expressed by means of 'will'), pronouncements, excuses and commitments (which don't seem to work properly without the directness of the indicative).[22] Admittedly, one could quite plausibly argue that you *can* express a promise by means of a subjunctive, as in

(99) If you washed the dishes, I would do the laundry/would be very grateful,

but I have mixed intuitions about these examples. At most, I think they may amount to indirect speech acts, in which the conventional illocutionary force of a sentence is used to make a different, but related contribution. Telling you what would happen if you did certain things entitles you to expect a certain behaviour from me, whereby I indirectly become conditionally committed to such a course of action.

My hypothesis is, then, that subjunctive conditionals are conventionally used to perform *simple* and not *conditional* assertions. Subjunctive consequents are inseparable from their antecedents, and can thus introduce no independent speech-act which in turn could be restricted by the subordinate clause. I shall take up the topic of subjunctives and their relation to indicatives in the following chapter—I shall argue that subjunctives are used to make claims about the world that are directly related to the acceptability of genuine conditional predictions.

Returning to our exploration of the functions of subjunctives, it is beyond doubt that there are meaningful questions featuring a subjunctive conditional clause:

(100) What would have happened if Nixon had pressed the red button?

[22] This phenomenon is quite stable across a range of different Indo-European languages. Regardless of the tense used in the antecedent of a subjunctive, most Romance, Germanic and Slavic languages use a special structure in the main clause, which has severely restricted uses beyond subjunctive conditional contexts. An exception would be Dutch, which can use simple past both in antecedent and consequent to express subjunctives (this us is sometimes found in old-fashioned English as well); however, there is also another option more akin to the structure used in English and note that neither can simple past accomplish many of the non-assertive tasks available to conditionals with indicative antecedents. I've also been told that Chinese doesn't mark the distinction between subjunctives and indicatives syntactically at all, and the intended interpretation must be worked out pragmatically. Unfortunately, I am not in a position to comment on this.

This, together with the fact that subjunctive antecedents cannot in general be used to restrict the scope of otherwise autonomous speech acts, further reinforces my suspicion that most of what might appear as conditional questions are simply questions about conditional assertions.

6.2. Restriction through indicatives

In the following two sections I will try to explain why we are better off keeping a thin meaning for indicative conditionals, disappointing though such a choice may be from the point of view of explanatory power. My suggestion is not that such a modest theory of meaning can account for all the puzzles indicatives face us with, nor is my intention to shrug off as irrelevant the puzzles others have worked so hard to solve. On the contrary, I would like to reinterpret the insights of Adams's and Kratzer's theory as pertaining not to the meaning of conditionals itself, but rather to the way we put this meaning into work in different contexts. I will show that this approach has several non-negligible advantages.

I wish I had more to say about the function of conditionals in restricting natural language quantifiers. I think it is an important topic that can shed a lot of light on how indicative conditionals work in natural language. And I have to admit that I do not have a seamless account of how their function in restricting complete speech acts, which I deem basic, makes them also apt for restricting some quantifiers within assertions themselves. However, I will take advantage of this opportunity to make some half-baked speculations.

6.2.1. Modals

We have seen that Kratzer's construal of bare indicative conditionals as implicitly modalised faces several difficulties.[23] The examples we have inspected suggest that modals perform the very same job in conditional consequents as in non-conditional sentences. It is therefore not necessary to modify the above minimal theory of indicatives to account for modalised consequents, as long as the modals involved admit of non-conditional uses. What is more, the theory has no difficulty explaining those bare conditionals that resist explicit modalising. It also allows us to give a straightforward account of the so-called biscuit conditionals, as we shall

[23] See page 119.

see in the following section. It is not clear how Kratzer's theory can rise to this challenge.

I for one do not think it is necessary or particularly enlightening to conceive of modals as quantifiers. It can be illustrative in some respects, as when 'it is possible that ψ' is rendered as 'there exists a possibility involving ψ', but I submit that it is rather an abstract reconstruction than a precise account of the message these modals convey. If you disagree, just think of how often you have as much as a hazy idea of the relevant range of possibilities when you utter a modal. Or think of the fact that usually, for you to utter 'may ψ' properly, ψ must be a distinct and relevant possibility in the context—this requires the corresponding range of possibilities to vary constantly and we seem to have no sound and independent grip on these variations. Last but not least, some uses of modals (like that of 'can' in 'he can swim') are rather difficult to reconstruct in this manner.[24]

The question remains as to how we should treat subjunctives—this will be the topic of the next chapter. 'Would' is a special expression, which requires at least an indirectly mentioned subjunctive conditional clause in order to work properly. Unlike with other modals, a consequent featuring 'would' cannot usually be used to perform an independent speech act and thus cannot normally be constrained by an indicative antecedent.[25]

Again, I don't think that treating 'would' as a quantifier is more than a matter of technical convenience. Its alleged duality with 'might' may give some independent plausibility to such a move[26], but I remain sceptical about this possibility in view of Stalnaker's observation that

(101) $A \diamond\!\!\rightarrow C$

is usually felt as equivalent to

(102) It might be that $A \mathrel{\Box\!\!\rightarrow} C$.[27]

[24] I have become wary of semantic analyses placing a lot of weight on modals—the latter accomplish too many different tasks in natural language to allow a straightforward uniform treatment. I have come to surmise that our notion of a 'modal' is primarily syntactic in nature. However, as my background in linguistics is sorely lacking, these are little more than unstudied hunches.

[25] There are of course exceptions, such as 'If you have a glass of juice at hand, I would appreciate it.' However, I think here the consequent is used to perform an indirect speech act, and as such, it doesn't amount to a counterexample to the above contention.

[26] See Section 4.2.2.

[27] (Stalnaker, 1981), p.99. However, I don't think this shows that the scope of 'might' is actually the whole conditional, as Stalnaker contends. I am only interested in arguing that 'might' performs by and large the same function in counterfactuals as in non-conditional statements, and that the fact that you can sometimes do without 'would' when you use 'might' is a mere quirk peculiar to English.

Moreover, in many Slavic languages, such as Russian, Serbocroat or Slovak, but also, for instance, in Dutch) this is the default and only way of expressing might-counterfactuals—you must use an adverb such as 'perhaps'. In Spanish, you can use a form of '*poder*' (may) analogous to 'might have + participle', but in a tense that is used to translate 'would+verb' constructions in English ('*podría haber* + *participio*').

6.2.2. Lewis's quantifier restriction

Once we have taken care of modals (basically by noting that we do not have to take care of them), it remains to explain Lewis's original insight about the role of conditionals in restricting quantifier domains. Recall that it concerned examples such as

(70) Always, if a man owns a donkey, he beats it now and then.[28]

This is what Lewis had to say about the use of conditionals in propositions such as (70):

> What is the price for the restriction-marking *if* to be a sentential connective after all? Exorbitant: it can be done if (1) we use a third truth value, (2) we adopt a far-fetched interpretation of the connective *if*, and (3) we impose an additional permanent restriction on the admissible cases. Let *If* Ψ, Φ have the same truth value as Φ if Ψ is true, and let it be third-valued if Ψ is false or third-valued. Let a case be admissible only if it makes the modified sentence either true or false, rather than third-valued. Then (39) is equivalent to (38) for all our adverbs, as desired [...].
>
> (Lewis, 1975), p.15

Lewis does his best to make these three requirements seem overly dramatic, but a closer look reveals that they follow quite naturally from Quine's interpretation of indicatives as conditional assertions. The 'far-fetched interpretation of the connective *if*' is simply the one we have opted for in this chapter.

I think this entitles us to surmise that in cases of quantifier restriction by conditionals, the scope of the quantifier is simply the whole conditional. An adverb of universal quantification, such as 'always', applied to a conditional, guarantees that anything that satisfies the antecedent will also satisfy the consequent. 'Sometimes', on the other hand, is used to make

[28] See page 115.

known that some of the things that satisfy the antecedent also satisfy the consequent.

As for vacuous cases, I wouldn't worry too much about them. They are very important in logic and mathematics, but much less so in real life. Our intuitions about them are not clear—and they can't be, since a vacuous quantified conditional, despite its syntactic complexity, makes no evaluable assertion to begin with. There is thus no point in using it and as a result, we don't. This is exactly the reason why most competent speakers of English unacquainted with first-order logic will be amazed to learn that 'All *A*s are *B*s' doesn't 'really' imply 'Some *A*s are *B*s'. Actually, the fact that our theory predicts the awkwardness of vacuous uses of quantified indicatives should count as a virtue rather than a drawback.

All this is admittedly very tentative and needs to be complemented by a deeper understanding of natural language quantifiers, which I unfortunately lack. Still, in view of the results I have briefly sketched here, I think some of the preliminary intuitions shoring up a conservative approach deserve at least the benefit of the doubt. My principal aim in this section was negative—to show that the language phenomena described by Lewis and Kratzer do not directly threaten the minimal theory of indicative conditionals I have advanced in the foregoing. We seem to be able to stick to a uniform treatment of indicatives across the wide range of their uses, which is in itself a worthwhile achievement. In the rest of this chapter, I will try to argue that the minimal theory actually fares better on a number of issues than an interpretation of Adams's semantics that has become standard.

6.3. Probability and indicatives

I have claimed that abstract systems of logic as a rule trade off generality for precision in modelling a specific aspect of the inferential behaviour of a piece of vocabulary, and I believe that Adams's probability logic is a case in point. It is extremely useful as an account of why certain valid inferences involving material conditionals sound paradoxical when translated by ordinary ifs. However, it leads up a blind alley when conceived as an account of the latter's meaning.[29]

The reasoning behind such a construal seems to be as follows: since

[29]It is not clear that Adams himself favoured this interpretation—I am inclined to think not, as he does not tend to express the purport of his theory in this way and avoids becoming entangled in the philosophical confusions to which it gives raise. I think it is in this connection that B. Skyrms dubs Adams's theory 'pragmatic', rather than 'semantic', although he doesn't elaborate much on the distinction: (Skyrms, 1994), p.13.

the truth-functional semantics for the material conditional gives rise to paradoxes, the ordinary 'if' cannot be material. This means that it must possess a different semantics, which yields 'the right logic'. The probability interpretation of indicatives has independent motivation and, what is more, doesn't validate the controversial inference patterns. This makes it into a good candidate for elucidating the meaning of natural language indicatives.

The problem with this argument is that while it rejects the material interpretation of conditionals, it doesn't question the theoretical framework in which that interpretation is set. Surreptitiously, it exports to natural language analysis the way of talking about meaning suited primarily to the purposes of a formal calculus. But if the hypothesis of the preceding section is true, then the inferential patterns we observe in natural language are not the sole product of the meaning of the expressions involved—this is a mere simplification that makes formal calculi so handy to use. So despite the appearances, the job of the semantics of a formal system devised to model the behaviour of a natural language expression does not necessarily consist in specifying the meaning of the latter.

Possible world semantics provides a particularly telling example in this respect—it sometimes comes in useful in order to model certain features of modal talk perspicuously, but it does not offer any insight into the actual meaning of modals (that is, into the kind of function they perform in our conversational exchanges) unless one accepts an extreme realism about possible worlds and a descriptive construal of modal talk. Indeed, an inadequate conception of the job of formal semantics seems to be at the heart of such an outlandish philosophical position.[30]

Adams's semantics is, in my view, a good approximation to the dynamics of our conditional beliefs; in particular, I believe it gives a correct explanation of our intuitions about the paradoxes of material implication. But we should not get misled by the word 'semantics'—the many merits of Adams's theory do not compel us to draw the conclusion that indicative conditionals are mere devices for expressing subjunctive conditional probabilities. And that is just as well, as such a view is ridden with difficulties.

6.3.1. Subjectivity and truth

The first and foremost problem with taking probability semantics for indicatives too seriously is that the latter then seem to be unfettered by any constraints imposed by objective reality. If the meaning of an indicative is

[30] See Section 4.3.1.

simply 'My subjective conditional probability of the consequent given the antecedent is high', then no matter how outlandish your belief system is, your assertion of an indicative will be unassailable provided you describe it correctly. This, of course, is at loggerheads with our intuitions about conditional assertions—I don't have to suspect you have been insincere or confused about your own beliefs in order to put your indicative into question.

A traditional way to counter this problem is to resort to some kind of expressivism about indicatives. As we have seen, logical empiricists originally came up with semantic expressivism as a strategy to account for the meaning of those utterances that didn't seem to make claims about possible experience. They explained moral injunctions, aesthetic judgements and in general any assertion whose acceptability could not ultimately be decided by appealing to some sensory experience, such as the outcome of an experiment, as expressions of subject's internal states, such as wishes, preferences, plans or impulses. They were considered to be just a sophisticated alternative to screams, sighs and smiles, and as such they could not properly be labelled as either true or false—predicates reserved for descriptions of physical reality as it might appear to our senses.

Even though logical empiricists are few and far in between nowadays, there has been a proliferation of different sorts of expressivism, with often very different theoretical backgrounds. Most of them have blurred the divide between properly empirical or descriptive statements and those that have expressive content, accepting that truth and falsity can be predicated of all of them. Sometimes, an expressivist position seems to boil down to the contention that not all proper assertions are 'descriptive'.[31]

In this dissertation I have distinguished between 'unproblematic descriptions' and descriptions *simpliciter*. Unproblematic descriptions are direct translations of sensory input into basic descriptive vocabulary and only exist in a linguistic Switzerland.[32] However, once we realise that descriptions are hardly ever this straightforward, I think there is no impediment to attributing a descriptive import even to highly theoretical or normative vocabulary. It is quite natural to conceive statements such as 'X is interesting' or 'Person Y holds belief Z' as describing the environment or, if you like, 'the world'—I surmise that those who deny this have in fact something akin to the unproblematic notion of description in mind.

[31] See (Frápolli and Villanueva, 2012).

[32] A good example of this is the relation between states of affairs and elementary propositions in Wittgenstein's *Tractatus* (which, incidentally, deeply inspired logical positivists)—the immediacy of translation is guaranteed by the structural isomorphism between the two.

However, we shall see in the following chapter that even much more elementary claims about the world than those that have traditionally been consigned to the expressive realm are far from unproblematic.

To sum up, while many contemporary strands of expressivism are admittedly very subtle, I think that we should in principle be wary of an explanatory strategy originally necessitated by a conception of language as inadequate as that of logical positivists. We certainly do not need it in order to explain the meaning of indicative conditionals. When uttering a conditional, you simply are making a conditional assertion, an assertion subject to the antedecedent obtaining. Despite its indirectness, such an assertion amounts to a valid autonomous move in a language game and it can in turn be endorsed or rejected by other speakers, which is sufficient for truth or falsity. There are thus no mysteries about what indicatives mean and how they accomplish this task.

Now, of course, in order to properly use a conditional, just as with any other sentence, you should make sure that it is the right thing to assert. Here is where conditional probability enters the picture. To be certain about a conditional assertion is often to have confidence that the assertion will be true should it enter into force. High conditional probability of the consequent given the antecedent is therefore exactly what you need—that's why it was called *conditional* probability in the first place (long before disputes about semantics of conditionals saw the light of the day). Just as the fact that we make ordinary assertions based on the subjective probability that we award to the proposition thus expressed doesn't make those propositions subjective, the fact that we make them based on our subjective conditional probabilities doesn't make conditional assertions subjective.

As we have seen, if you make truth into the rock bottom of any semantic inquiry, a simplistic descriptive picture of language is often a natural consequence. Asked about what makes a particular utterance of a sentence have the truth value it does, there is, on this picture, not much more to answer than 'the way things are' or simply 'the world as it is'. Being true then means representing the reality faithfully. As a result, the fact that most indicative conditionals are only useful in situations of uncertainty, as Lewis's and Edgington's proofs show, invites the conclusion that they can almost never be properly said to be true or false.[33]

I submit we should pay more heed to our linguistic gut feelings and resist this temptation. Once you conceive of language as a device for hu-

[33]This is a conclusion that Bennett takes to enormous lengths—with a few careful exceptions, he never calls an indicative true or false throughout the entire (Bennett, 2003). What a remarkable and superfluous feat!

man communication rather than a bunch of syntactic rules together with a semantics that divides your sentences into two disjoint classes, truth itself will be in need of clarification. Truth is then not some kind of mysterious property of uttered sentences, but rather a device speakers use among other things to endorse assertions which they for some reason do not wish to or cannot repeat themselves.[34] The nature of these assertions is completely irrelevant—they can of course be descriptive, but they can also convey modalised statements, value judgements, precepts and instructions, advice, censure or, for that matter, indicative conditionals. The grounds upon which speakers may make these assertions are undoubtedly interesting, but do not seem to have much to do with the nature of truth.

Triviality results, *contra* a generalised interpretation, do not show that indicative conditionals are somehow non-factual, as opposed to regular, descriptive propositions. The only distinction these proofs warrant is between conditional and non-conditional sentences—the latter may express, as we have seen, all kinds of non-descriptive propositions. The fact that conditionals non-trivially increase the expressive power of our language with respect to non-conditional sentences is no good reason to deny truth-aptness to conditional assertions. There is no law to the effect that only assertional contents representable as sets of possible worlds deserve the lofty title of 'proposition'. It is people, not possible worlds, who attribute truth or falsity to assertions. In natural conversation, indicative conditionals are readily treated as true or false—if our theory forbids this, maybe we should change the theory instead of our, in this case utterly innocent, linguistic habits.

6.3.2. Biscuit and other conditionals

Even though many indicatives are asserted based on an inference whose propriety they in turn serve to sanction, which makes their association with conditional probabilities natural, this is by no means always the case. Austin came up with an example that has given a name to a whole group of such conditionals:

(103) There are biscuits on the sideboard if you want them.[35]

Note that even though this sentence sounds completely natural to our ears, it differs from most ordinary indicatives we utter in that the consequent

[34] This is the view advocated in (Frápolli, 2013).

[35] (Austin, 1970), p.210. Austin used this example to show that not all ifs were 'causal conditional', but did not further elaborate on the phenomenon.

is felt to be independent of the truth of the antecedent. The subordinate clause does not limit the truth of the main clause—typically, you will be deemed to have committed yourself to the presence of biscuits on the sideboard whether I want them or not. Neither does the antecedent play any role in inferring the consequent; the latter has usually been ascertained on independent grounds. It seems, rather, that the antecedent specifies the circumstances in which the contribution made by the consequent would somehow be to the point.

The foregoing suggests that invoking conditional probability cannot directly explain the assertibility of the above conditional—arguably, $p(C/A)$ is high, but this is only because $p(C)$ is high irrespective of the value of A and Gricean considerations would therefore strongly recommend asserting C instead of $A \rightarrow C$. However, this doesn't square with our intuitions about (103)—in a typical context of utterance, it would be rather more appropriate than mere

(104) There are biscuits on the table.

How come? If conditional probability is all there is to indicative conditionals, there is no reason to go for a whole conditional instead of its consequent.[36] You may of course try to explain (103) and its kin as deviant conditionals that acquire their standard interpretation by means of some kind of pragmatic adjustment, such as conversational implicature. But this contradicts the spontaneity with which we utter biscuit conditionals—we don't seem to perceive any tension in such utterances, as we probably should if they were to trigger a conversational implicature. Moreover, it is not clear how you could calculate something like a 'contextual relevance interpretation' out of the meagre semantic input provided by the probability calculus.

According to M. Franke, there are actually two questions to be answered in relation to biscuit conditionals : (i) how their independence interpretation arises, and (ii) what their discursive function is.[37]. I think the answer to the first one has to do with shared world knowledge. Both you and I know that your wanting the biscuits cannot by any means bring about their presence on the sideboard, so it is safe for me to conclude that you have independent grounds to assert the consequent. Furthermore,

[36]The difference with 'even-if' conditionals is that the latter are properly uttered in a situation where the consequent is by and large accepted and therefore assertible, but there remains a doubt as to whether it will still hold in the (extreme) circumstances specified by the antecedent. With typical utterances of biscuit conditionals, however, there is no such residual doubt—the independence between the antecedent and the consequent is total.

[37](Franke, 2007), p.92

different background assumptions can trigger different interpretations of the same conditional:

(105) If you need me, I'll be there.

is ambiguous between a standard and a 'biscuit' interpretation. Your conditional can be interpreted as informing me of the place where I can find you in case I need help (and which obviously does not depend on my needing help or not), but also as a conditional pledge to lend me a hand whenever I cannot fend for myself on my own.[38]

As for the second question, we have seen that in (103), the antecedent specifies the condition under which the utterance of the consequent is relevant, rather than true. But there are other possibilities as well:

(106) The euro is an economic basket case, if you ask me.

(107) If I may say so, your wife has a very strange sense of humour.

In these cases, the function of the antecedent consists in hedging the assertion of the consequent against possible social inappropriateness.

The question is, should we postulate a different meaning for each of the above uses of conditional clauses (introducing an epistemic, relevance, and social acceptability restriction)? At this point it should be obvious that the answer is negative. With our minimal semantics for indicative conditionals, we can straightforwardly account for this usage. Just as you can use the same hammer to bang in nails and crush coconuts, you can condition your assertions to achieve a number of different discursive effects. Perhaps most of the times it is because you cannot guarantee that your assertion will be true unless the condition holds, but there are other possibilities as well.

In (103), your assertion is conditional upon my wanting the biscuits, in which way you manage to convey information about the intended purport of your assertion. To understand better what is going on, think of the following two exchanges:

Dialogue 1

[38] DeRose and Grandy sometimes speak as if it these implicatures were due to different speaker's motivations (relevance and probability) in making the assertion: (DeRose and Grandy, 1999), pp.413-414. I think the right thing to say is rather that it is the presence of these contextual implicatures which triggers the biscuit or epistemic reading of the conditional assertion in question. This is just splitting hairs, though—I otherwise agree that biscuit conditionals provide the best evidence for speech-act-based theories of indicatives (but see footnote 42).

You: Do you want any biscuits?
Me: Well, not really—I am a coeliac. But thanks for asking.
You: Don't mention it.

Dialogue 2

You: Do you want any biscuits?
Me: Actually I do. I've grown rather hungry after all the hula-hooping.
You: There are some on the sideboard.
Me: Thanks.
You: Don't mention it.

Observe that in both dialogues, the question that opens the exchange can be interpreted as making an offer rather than merely asking for information, but this is accidental. The main point is that whether you make your assertion about the biscuits or not depends on my answer to that question—if I don't want them, there is no point in drawing my attention to their location.

The simple contention I want to put forward here is that a biscuit conditional streamlines the exchange by avoiding the intermediary step of my explicit answer to your question by transforming the latter into the antecedent of a conditional. And it can accomplish this task simply in virtue of the standard meaning of indicative conditionals I have explored above—there is no need to posit any deviant interpretation of the if-clause. The different flavours in which biscuit conditionals come have to do more with the directionality of the conversation at hand than with the actual meaning of indicatives.[39] Far from being 'jejune'[40], the topic of biscuit

[39]Franke comes to a similar conclusion in (Franke, 2007), but he is in my view overly optimistic about the prospects of explaining biscuit conditionals with the exclusive recourse to standard dynamic semantics. If the only purpose of an utterance is to effect a change in a context set represented by a set of possible worlds, then in case of (103), updating with the consequent is of course tantamount, or even preferable, to updating with the whole conditional. It is not clear how contexts thus construed can codify information about 'optimality of utterance' to remedy this difficulty. I think that to bestow more plausibility on his approach, Franke implicitly refers to the intuitions *underlying* and *justifying* the dynamic semantics (such as those we have presented here), not the formal structure itself. In particular, this structure cannot explain utterances of conditionals as conditional utterances.

[40](Bennett, 2003), p.126

conditionals shows the superiority of a lean account of this meaning.

Finally, while biscuit uses of indicatives are extremely common, as far as I know nobody has come up with a plausible example of a subjunctive biscuit conditional. Mackie suggested

(108) He was there all the time, had I but known.[41]

but it doesn't strike me as very convincing. It doesn't seem to admit of standard paraphrases—it is much better rendered as

(109) If only had I known that he was there all the time!

This further shores up the conjecture, made in Section 6.1.3, to the effect that subjunctives are used to make *simple* rather than *conditional* assertions.[42]

6.3.3. The futile search for the 'right' logic

We have already seen that if you construe indicative conditionals exclusively as vehicles for conveying information about subjective conditional probabilities, you will be tempted to posit polysemy each time you encounter a use resistant to such explanation. This happened with biscuit conditionals, and this happens again when it comes to the material conditional. In my view, perhaps the most egregious claim that the conflation of formal and natural languages in modern philosophy of language has engendered is the claim that in the context of mathematics, our ordinary indicative conditionals all of a sudden change their meaning and become truth-conditional. So it would seem that we sometimes utter our conditionals as Adams's conditionals, sometimes as material conditionals and sometimes as relevance or biscuit conditionals. On this view, to understand a speaker's intent in uttering a conditional, you first have to resolve the vagueness that underlies natural language ifs.

[41] (Mackie, 1973), p.73

[42] Even though everything seems to be falling into place, to be perfectly frank with you, I must admit to some residual doubts about whether this is all there is to be said about biscuits—there are several grammatical conjunctions apart from if that can switch between relating contents and speech acts. Compare 'Since you were so loyal to me, I haven't revealed your hideout to the police' and 'It was nice of you to accompany me, but you needn't have done it' with 'Since you were so loyal to me, here are the keys to my bedroom' and 'It was nice of you to accompany me, but now just leave me alone!' So maybe we should make the distinction between conditioning contents on the one hand and speech-acts on the other also for indicatives conditionals. I suspect, however, that this distinction would make little difference in our case—C conditioned upon A would count as asserted with A turning out to be true, and there would be no asserted content, and hence no assertion, in case A came out false.

I think that this is either a very infelicitous way to express a platitude (every uttered expression has to be interpreted in a pertinent context if we want to grasp the speaker's purport), or a token of a deeply flawed semantic methodology. Unfortunately, most scholars seem to lean towards the latter option. Needless to say, one of the main goals of this dissertation is to show the unreasonableness of gratuitous ascriptions of polysemy to indicatives.

To think that our conditionals possess different linguistic meanings on different occasions of use conflicts directly with Grice's Modified Occam's Razor[43], which sums up some of the most fundamental intuitions guiding semantic analysis. This position, moreover, sometimes requires attributing proficiency in manipulating meanings with which a person may be completely unfamiliar (in the case of the material conditional). How come that pre-Fregean mathematicians used conditionals in their work without feeling the urge to warn the reader that their meaning differed substantially from that of vulgar ifs?

Those who attribute different meanings to what seems to be a unitary expression have to shoulder the burden of explaining how the homonymy they posit comes about. In extreme cases, it will be down to pure accident, as when originally distinct roots evolve into the same linguistic form. However, this can't be the case with indicatives, whose divergent uses all seem to preserve some crucial trappings of conditionality. Our task is thus to explain what it is about 'ifs' that makes them naturally fit to play all these roles.

We ordinarily tend to chalk up divergent, but acceptable uses of our vocabulary to contextual pressures triggering a range of pragmatic effects over a common semantic core. Why, then, should we treat the case of conditionals any different? I think the answer again has to do with the flawed theoretical background from which we approach the problem. Since to put forward a theory of conditionals has come to mean tp devise a formal deductive system with a suitable semantics, if two different systems are equally useful in different contexts, the recourse to polysemy seems as the obvious way to go.

Nevertheless, this temptation should be resisted lest we end up riding roughshod over the MOR. Remember that formal semantics, despite its name, does not have to be construed as establishing the meaning of the expression under study. It suffices if the semantics models some features of its use in a useful and interesting way. These features may be motivated by its meaning, but they may also be due to accidental properties of a

[43]See Section 1.3.1.

prominent context of use. I believe this is exactly the case with the material interpretation of indicative conditionals.

Mathematicians had been using conditionals in their expositions long before Frege came up with his ingenious conditional stroke. This technical device (and Frege never doubted that it was one), despite raising many doubts as to whether it was an adequate counterpart for 'if', proved miraculously useful in codifying conditionals in mathematics. Hence the natural idea that this simple truth-functional account, bolstered by a few pragmatic considerations, might encompass all uses of 'if' in our language.

We have by now seen this theoretical hunch to be unsustainable. However, once they recognised the inviability of the project of reducing all ifs to material conditionals, most scholars only switched their bets to another horse, while keeping the reductionist mindset. Instead of returning to the point of departure, where the material conditional was a mere formal contrivance designed to emulate some of the functions of natural indicatives, they asked: if indicatives are not material conditionals, what are they? Or in other words: if the truth-functional semantics doesn't capture the meaning of ifs, which semantics does? And since probability-based semantics fared far better on most counts (and for a good reason, as we shall presently see) than the truth-functional one, it was handed the palm by all but a few eccentric champions of the material conditional.

Let me repeat myself, just to drive this important point home. Application of probability to the study of how people use conditionals is more than welcome. The crux of the problem lies elsewhere—it is the assumption that formal semantics, including Adams's, always spells out the core meaning of the modelled expression. That's why we have run into problems such as those I have surveyed in this section—subjectivity, truth and biscuit conditionals. And here comes another—if we assume that the only job of indicatives is to express subjective conditional probabilities, then the use of conditionals in mathematics, where the recourse to probability appears out of place, becomes a conundrum again.[44]

With the minimal account of meaning championed here, there is no reason for such worries. Conditionals perform the same function in mathematics as everywhere else. However, the peculiar context of mathematics has a distinct impact on how this function translates into actual inferential behaviour. This, I submit, is how we should go about explaining the surprising usefulness of the material conditional in formalising mathematical reasoning.

Ironically, in order to explain in detail why indicatives behave like ma-

[44]See (Suppes, 1994), p.5.

terial conditionals in certain contexts, it is extremely helpful to look at the positive arguments marshalled by the diehard supporters of the identity thesis. If we can pinpoint the precise juncture at which they go awry, we can perhaps identify the assumptions that underlie a material interpretation of indicatives and show that they typically accompany mathematical reasoning, but may fail in other contexts.

The arguments by W.H. Hanson in (Hanson, 1991) are a good case in point. In itself, Hanson's work shows a crass disregard for the niceties of the debate about the meaning of indicatives and doesn't even attempt to address the many compelling objections that have been raised against the identity thesis. The argument is surprisingly simple-minded: with the help of conditional proof (deduction theorem), you can derive a conditional from a disjunction; since the other direction is uncontroversial and the conditional proof is valid, conditionals are equivalent to disjunctions—just as the identity thesis maintained the whole time.[45] Naïve as it is, the very simplicity of this argument allows us to put our finger on the precise spot where the material interpretation of indicative conditionals kicks in.

Of course, the real question, shrugged off by Hanson, is whether reasoning by conditional proof should be regarded as valid for indicatives in the first place. We have seen that it is *not* valid for Adams's conditionals.[46] This is because conditional antecedents are treated as new information that can wreak havoc among the previously accepted premises. Apparently oblivious to Adams's arguments, Hanson contends that

> For the antecedent of the indicative conditional "If A, then B" asks us to suppose that (in addition to whatever else we may believe) it is in fact the case that "A". It thus seems appropriate that cp for indicative conditionals merely adds "A" to the stock of sentences that have already been assumed or derived.[47]
>
> (Hanson, 1991), p.59

So how do antecedents of conditionals work in real life? Can the minimal theory of meaning (if it deserves such a name at all) I have advanced here decide this question?

---

[45] An analogous argument, with exactly the same weak spots, is presented in (Rieger, 2013), pp.3163-3167.

[46] See page 38.

[47] Compare (Rieger, 2013), p.3164, on $A \models B \rightarrow C$: 'What is distinctive about the indicative conditional is that, in making the assumption of B, we leave A undisturbed. The conditional concerns how things are, not how they might be, and the supposition of B plays the role of additional information about the world.'

The answer is that it cannot and it is good that it cannot, because the impact of an antecedent on the rest of our beliefs is not a matter of meaning of the indicative conditional involved, but is rather determined by the entire inferential context within which the conditional is asserted or evaluated. Once again, in making a conditional assertion, you commit yourself to the consequent in case the antecedent turns out to be true. In a normal setting, you are aware that certain eventualities would make you change your mind about what is actually the case, and the use of indicative conditionals naturally reflects this fact. Since Adams's probability logic is designed to take the possibility of precisely such changes of mind into account, it is in a good position to explain why inference patterns such as CAC and NAAC sound paradoxical.[48]

However, sometimes your antecedent will be independent of your other beliefs and as such will function just as an additional premise. Or you may legitimately decide that you are going to keep some of your premises fixed, in order, say, to explore their inferential power. This can be done implicitly, or explicitly, for instance in the form of an axiomatic system. In such circumstances, the antecedent of an indicative conditional *will* be correctly construed as asking you to suppose that it is true *in addition to* whatever else you may believe, just as Hanson would have it. It is therefore no coincidence at all that material conditionals made their entrance in the philosophy of language together with the first serious attempts at axiomatising mathematics—as in a monotonic system you can derive a conditional from a disjunction and vice-versa, you can simply substitute the former for the latter.

That this fixing of the premises often occurs implicitly and sometimes even unconsciously is witnessed by the fact that inference patterns such as contraposition, hypothetical syllogism and antecedent strengthening (that all have counterexamples in natural reasoning) have been regarded as valid throughout the history of logic, while there was no question of justifying them by recourse to a truth-functional semantics for the conditionals involved. This is because it is often very useful to work in a monotonic inferential system, where your axioms and auxiliary premises are temporarily sheltered from doubt, in order to see how far you can go without having to shed them. Once you find the consequences of your axioms unpalatable, you will be compelled to abandon the system as a whole and try your luck with another one.[49] In the meantime, however,

[48] See page 10.

[49] The classical requirement that axioms be 'self-evident' is independent from the axiomatic method itself and has been regarded as superseded at least since the advent of

the idealisation presupposed by keeping the axioms unassailable greatly streamlines the inferential process by justifying the use of stronger rules of inference.

Ordinary reasoning, on the other hand, has to adapt dynamically to new information and cannot choose to protect beliefs from contrary evidence at will. This is because we use the information at our disposal chiefly to make decisions. We cannot afford to make wrong decisions 'for the sake of the argument', as their consequences will be here to stay even if we change our mind afterwards.

Adams's logic, by taking into account all Kolmogorov probability distributions, makes allowances for any rational belief change that could take place upon the impact of the new information conveyed by the antecedent of an indicative. In a sense, whereas in classical logic your assumptions represent the rock bottom of your inquiry, for Adams nothing that could possibly go wrong with the assumptions in view of an additional hypothesis should be glossed over. Ironically, its faithfulness to possible belief change makes Adams's logic purely negative with respect to the classical and as such largely useless in deductive sciences. This logic, just as many other indicative and counterfactual alternatives to classical logic, doesn't seem to have other function besides displaying the adequacy of the associated semantics by invalidating intuitively suspicious inference patterns. It is, however, never really used in interesting proofs, as is the case with classical logic in set theory and analysis. This explains the feeling that classical logical has a prescriptive flavour to it, whereas the alternative logics we have surveyed here seem for the most part descriptive; so that predictions at odds with shared linguistic intuitions are much more deleterious for the latter than for the former.

So does a disjunction imply the corresponding conditional after all? I think there is no clear-cut answer, since it depends on the phenomena you wish to model by means of the posited entailment relation. In formal languages, relations of logical consequence are exhaustively underwritten by the semantics of the involved vocabulary. We have seen, however, that as a rule, inferential relations in natural language are much more convoluted, depending as they do on a plethora of contextual factors. There is no one correct entailment relation for natural language, waiting to be discovered by a particularly shrewd philosophical logician.[50]

Take any indicative conditional $A \rightarrow C$ about the future, say

non-euclidean geometry. Frege, among others, had to learn this the hard way.

[50]The arguments by J. Etchemendy in (Etchemendy, 1990) can perhaps be interpreted as pointing in the same direction even for a more formal domain of inquiry.

(110) If you work hard enough, you'll finish your PhD on time.

Suppose you're a new graduate student, you are a little overwhelmed by the academic environment, and I say (110) to reassure you about your chances of success. However, as it happens, two months into your PhD, your supervisor falls in love with a circus performer, loses all his fervent interest in non-Russellian semantics for definite descriptions and leaves academia to travel around the world in a caravan. You are left stranded in a philosophy department whose all other members are too busy untangling Deleuze's interpretation of quantum mechanics. Despite all your efforts, your grad school period is an unmitigated disaster and you decide to spend the rest of your life cultivating a garden.

Seen in retrospect (while you're digging up some beets, for instance), my utterance of (110) cannot appear more misguided. It was obviously false. Whenever A turns out true and C false, the assertion of $A \to C$ can safely be labelled as such, no matter how improbable the events that have made the inference justifying the conditional fail were at the moment of utterance. Once you make a public assertion, you lose control over it—in particular, you cannot retract it if things take an unexpected turn. In a sense, then, committing yourself to $A \to C$ means committing yourself to $(A \wedge B) \to C$ for whichever B might occur; otherwise we couldn't evaluate your assertion as false in a situation where A, B and $\neg C$ all hold.

However, this of course does not mean that if I believe in (110), I should by the same token believe in

(111) If you work hard enough and your supervisor falls in love with a circus performer, loses all his interest in your topic and leaves the academia, you will finish your PhD on time.

What really happens is that in asserting (110), I bet that all those circumstances (most of which I cannot as much as envisage) that would defeat the inference shoring up my conditional prediction will not in fact occur.[51] I can therefore both believe $A \to C$ and disbelieve $(A \wedge B) \to C$ without being irrational. This implies that at the level of beliefs, they are compatible.

My contention is simply that different viable conditional logics model different inferential relations between conditional assertions. There is no unique correct logic of indicatives, and neither does deviant inferential behaviour of some indicatives in certain contexts justify us in attributing

[51] In the next chapter, I will say that you bet that the events will follow a particular *script*. Since subjunctives are primarily used to explore such scripts, their inferential behaviour is analogous to this 'logic of conditional beliefs'. This, I submit, is the link between indicatives and subjunctives.

a different meaning to them. Again, in classical logic, premises are treated as unassailable, so conditionals can easily be construed as monotonic with respect to the strength of their antecedent. This is useful in modelling situations where adopted commitments are for some reason not retractable, such as those explored in axiomatic systems. Non-monotonic logics, on the other hand, are good at modelling the principles according to which individual people award credence to conditional assertions.

I think these considerations also explain the clashing intuitions people have about the validity of SDA, at least for indicative conditionals. If I assert that

(112) If you work hard enough or plagiarise some obscure German treatise, you will finish your PhD on time,

I may do so only because I am convinced that out of the two possibilities mentioned in the antecedent, the first one is by far the most likely and it does guarantee your finishing on time. However, should you against all odds secretly take evening German classes, spend a summer burrowing into some Bavarian philosophy library and get flunked at the defence by a committee suspicious of your predilection for interminable sentences and mind-boggling agglutination of nouns, adducing the reasons I might have had for asserting (112) will be to little avail. My assertion will be irremediably false.[52]

Here intuitions seem to be even clearer than in the previous case, because my explicit mention of the second disjunct in the antecedent of (112) is felt to commit me quite strongly to the conditional

(113) If you plagiarise some obscure German treatise, you will finish your PhD on time,

as I'm assumed to have given some thought to such a possibility (this can certainly be explained on the strength of some Gricean principles). But overall, the moral is very much the same as before—once your conditional is publicly asserted, you lose control over it and any way that the antecedent may come out true is relevant for evaluating it. This is also the reason why simplification of disjunctive antecedents is so intrinsically linked to antecedent strengthening—replacement of equivalent antecedents only plays an intermediary role on this picture, by making it possible to spell out the tacitly undertaken commitments. This explains

[52] M. Woods also argues that SDA poses a problem for accounts entailing that assertibility of indicatives is exhausted by corresponding conditional probability: (Woods, 1997), p.27.

the somewhat surprising fact that classical logic with the material conditional actually fares much better with respect to our intuitions about SDA than most of the alternatives.

The above phenomenon can be exploited by speakers in order to explicitly convey that they do not regard certain options as live possibilities. It has frequently been observed that SDA seems to fail in examples such as the following:

(114) If the United States devotes more than half of its national budget to defence or to education, it will devote more than half of its national budget to defence.[53]

(114), the argument goes, cannot be taken to imply

(115) If the United States devotes more than half of its national budget to education, it will devote more than half of its national budget to defence,

as it is plainly unacceptable. However, this might be the whole point of asserting it—to signal that the speaker considers the possibility of the US earmarking more than half of its budget for schools to be far-fetched. Compare

(116) If the United States devotes more than half of its national budget to education, then pigs have wings.

Here, the speaker doesn't mislead us at all by using a disjunct she has already ruled-out, as her utterance makes this perfectly clear.[54] All in all, by shifting the focus on the actual assertive use of indicatives instead of treating them as mere vehicles for expressing conditional probabilities, we can shed a lot of light on what is at stake in the debates about their inferential properties.

[53]Based on (Nute and Cross, 2001), p.29

[54]Note that this problem doesn't only arise with disjunctions, but existential quantification as well: 'If the US devotes more than half of its national budget to anything, it will be defense.' For a discussion of SDA in the context of subjunctives, much of which, however, applies equally to indicatives, see page 76.

Chapter 7

Indicative and subjunctive conditionals

Up to now, I have for the most part been busy showing how many of the apparent semantic conundrums indicative conditionals face us with are due to an excess, rather than a lack of philosophical theorising. It turned out that we had all the ingredients for a viable account on the table—we only had to shed some philosophical prejudice and combine them correctly in order for the global picture to fall into place. It sufficed to put down the sledgehammer and go for a nutcracker instead.

I believe the situation is different with subjunctives. We use them every day in all kinds of situations, but our theoretical understanding of how they work is sorely deficient. They are assertions about unrealised possibilities and as such challenge our natural tendency to revert to a simplistic view of language as a mirror of the world (as if 'the world' were a completely unproblematic notion). I think that the fact that we still lack a full-fledged theoretical alternative to a Tractatus-like conception of language is the main reason why so many aspects of our use of subjunctives remain shrouded in mystery.

How can this be, given the exorbitant amount of philosophical research into this topic since the 1950s? There has undoubtedly been some progress, at the very least in ruling out inviable theoretical options, but it has been painfully slow. All too many years were spent talking about the formal properties of an unfathomable similarity relation, which rather than throw light upon subjunctives made them seem even more arbitrary and purposeless. Given our methodological decision to treat language as a tool kit and to explain the meaning of individual expressions in terms of the role they are conventionally fitted to play in speech acts, such a theory must at this point inevitably appear suspicious. Leaving Gauker's and

Lycan's largely vacuous theories aside, neither did Kratzer's attempt to merge insights from both cotenability and similarity theories prosper. We have seen that in the end it was the causal theories that seemed to have done the best explanatory job; however, it was at the cost of making the formal semantic machinery practically redundant.

In this chapter I'll make a rather speculative attempt at explaining the relation between indicatives and subjunctives, combining insights inspiring both causality theories and the Adams-Dudman-Edgington approach based on the analogy between subjunctives and future indicatives. I think that this analysis suggests that an adequate and interesting treatment of subjunctives requires a profound change in how we look at both language and human cognition. I will not be in a position to give all the details of a full blown alternative theory, but I will try at least to point out the way I deem worthy of exploring.

7.1. Subjunctives as past predictions

Since the very beginnings of the analytic inquiry into conditionals, scholars have searched for a theory that could bridge the gap between indicatives and subjunctives. Cotenability theorists, as we have seen, thought that the distinction did not warrant any special attention—according to them, the subjunctive form was preferred when the antecedent was disbelieved, but there was no question of two different conditional functions. It was not until Adams came up with his pair of Oswald conditionals that philosophers awoke to the more fundamental differences between them.[1] Since then, however, another extreme position has come to hold sway in the literature—given the lack of overlap in the most popular semantics for the two constructions, they are regarded as fundamentally different and the very project of looking for connections between them is rejected as misguided.[2]

We should avoid both extremes. Adams's pair shows that indicatives and subjunctives convey different contents; but there are numerous parallelisms as well.

[1] See page 6.

[2] I think D. Lewis, hypostasing formal semantics while defending completely divergent ones for indicatives and subjunctives, bears a lot of blame for the success of such a view. Usually, this theoretical position is adopted rather tacitly (take for instance Bennett, whose account of the link between the two types boils down to a few casual remarks about the formal structures behind their logic: (Bennett, 2003), pp.174-175), but an explicit avowal can be found in (Gibbard, 1981).

First, there are of course syntactic analogies. To express both indicative and subjunctive conditional contents, we make use of largely overlapping linguistic structures. The most obvious example is the grammatical conjunction 'if', but a plethora of others ('on condition that', 'in the event that' 'provided that', etc.) work just as well for both of them. Furthermore, in English, a typical subjunctive conditional can be obtained from a future indicative by means of a past tense shift—'does-wills' become 'did-woulds', which can further be transformed into 'haddone-wouldhaves'.[3]

Syntactic links between indicatives and subjunctives come accompanied with similarities at the level of content. In particular, the important symmetries between

(I) future indicatives and subjunctives, and

(II) future indicatives and past subjunctives uttered at a later point

also suggest that there is a lot these two constructions have in common.

The first point is illustrated by the fact that indicatives and subjunctives about the future are felt to make basically the same assertion (there is a slight preference for the subjunctive if the antecedent is deemed unlikely, but this does not substantially affect its interchangeability with the corresponding indicative):

(117) If Russia continues its aggression in Ukraine, the EU will consider sending another protest note.

(118) If Russia continued its aggression in Ukraine, the EU would consider sending another protest note.

The second point is witnessed by the fact that someone who on November 21, 1963 claims that

(119) If Oswald does not kill Kennedy, no one else will.

could exactly for the same reasons have claimed on November 23, 1963 that

(120) If Oswald had not killed Kennedy, no one else would have.

These and analogous observations have led some theorists to conceive of counterfactuals as expressing some kind of past tense of future indicatives. I shall present two such theories: E. Adams's epistemic past tense

[3]We shall see that whereas the first point holds across many different European languages, the second one does not.

construal of subjunctives and V. Dudman's relocation theory. Afterwards, I shall systematically explore the scenarios in which the symmetry between does-wills and had-woulds fails, trying to glean a clear view of the kind of contribution counterfactuals really make. We shall see that this inquiry will naturally connect a (duly qualified) tense-shift account with the prominent role causal considerations are deemed to play in our evaluation of counterfactuals.

7.1.1. Adams's and Edgington's 'past probability'

As I have already noted, the position that has become quite standard in the literature (witness for instance J. Bennett or F. Jackson) combines probability semantics for indicatives and some kind of similarity semantics for subjunctives. We have seen good reasons to be suspicious about both of them; however, their combination produces yet another undesirable effect—an almost complete disconnection between the functions of the two sorts of conditionals. A theory is therefore called for that could account not only for their differences, but also for the many parallels.

E. Adams sketched a theory of subjunctives that would connect them with indicatives through a probabilistic interpretation of the latter. Even though ultimately unsuccessful, Adams's attempt (together with Dudman's theory analysed in the next section) will provide us with some resources for an amended account along similar lines. I shall argue, following Adams himself, that such an account will probably have to make use of the notion of (causal) dependence.

Probabilities of counterfactuals as prior conditional probabilities

The fundamental insight behind Adams's take on counterfactuals is rather straightforward—after affirming an indicative and finding out that its antecedent, consequent, or both happen to be false, we often forfeit the right to the indicative itself, but can still affirm the corresponding subjunctive. Subjunctive mood in conditionals can therefore be regarded as an 'epistemic past tense' of sorts.[4] Adams gives the following example:

> Two men are walking in the woods and spy a bird in the shadow in such a way that its colour cannot be made out. One man might use the indicative in telling the other "If that bird is a canary it will be yellow." Now, however, suppose that the bird flies out into the sunlight, where it is clearly seen to be blue and not yellow. Under

[4](Adams, 1975), p.103

> the circumstances the first man will be unlikely to continue to affirm the indicative [...]. On the other hand the first speaker will be likely to 'substitute the counterfactual for the indicative' and affirm "if that bird were a canary it would be yellow."
>
> (Adams, 1975), p.104

So it seems that the function of the subjunctive might consist in expressing conditional probabilities prior to ascertaining the falsity of the proposition conditionalised upon. Of course, since by no means does a neat story such as the one above invariably accompany our assertions of subjunctives, these prior probabilities may be merely hypothetical.

This construal, according to Adams, gives a hint about why in informal renderings of Modus Tollens inferences counterfactuals are strongly preferred to indicatives. The moment that the two men clearly see that the colour of the bird is blue, they can no longer endorse the indicative

(121) If that bird is a canary, it is (will be) yellow.

since not even the hypothesis that it is a canary can offset the obviousness of its colour. However, they can still affirm

(122) If that bird were (had been) a canary, it would (have) be(en) yellow.

Adams argues on Bayesian grounds that except for a few special cases, a person who awards high value to $p(B/A)$ should attach high probability to $\neg A$ upon learning $\neg B$, while lowering her value of $p(B/A)$ accordingly. That's why, on the past epistemic interpretation of the subjunctive mood, a subjunctive conditional can, whereas an indicative usually cannot, feature in a Modus Tollens.

Some problems

I think that just as in case of indicatives, while recourse to probability can undoubtedly shed light upon certain phenomena of the logic of subjunctives, it can be misleading if it is expected to give the whole story.[5] The acceptability of a counterfactual $A \mathrel{\Box\!\!\rightarrow} C$ for which a past assertion of a corresponding indicative $A \rightarrow C$ is conceivable seems to correlate more with the endorsement of some kind of inferential relationship between A and C in the circumstances than with the actual or hypothetical past acceptance of $A \rightarrow C$.

[5] Adams seems to have ended up admitting that much. He only addresses subjunctives in passing in his definitive (Adams, 1998), and is non-committal as to his stance towards the theory propounded in (Adams, 1975).

Take one of the special cases where Modus Tollens fails. Sometimes you may give credence to $A \rightarrow C$ only because you are already quite certain about C itself. Of course, in such circumstances you won't be inclined to conclude $\neg A$ after learning that C is not the case after all. However, and this is a crucial point, you cannot make out from a simple probability distribution that awards high values to both $p(C)$ and $p(C/A)$ whether your belief in C is *the* reason for believing $A \rightarrow C$.

A variation on an example by Adams will make this clearer. Suppose that before descrying the bird in the shadow, you believe for some reason that all birds in the area are yellow. Then your conditional probability of 'That bird is yellow' given 'That bird is a canary' will be high and you can certainly assert (121). What will happen once you see that the bird is blue? Obviously, you will no longer believe that all birds in the area are yellow. But what about the bird being a canary? There are at least two possibilities.

If your belief in the yellowness of all birds was your only reason for asserting (121), then you obviously lack any right to apply Modus Tollens after knowing better. However, if despite the failure of that belief you still sincerely accept that all or almost all canaries are yellow, to conclude that the bird is not a canary is still the right thing for you to do. Observe that while there is no way to distinguish the first scenario from the second attending exclusively to (conditional) probability values, the counterfactual (122) is supported if and only if the Modus Tollens inference goes through.

And, *pace* Adams, this doesn't occur only in special cases, since the dynamics of our beliefs doesn't always conform to the smooth model of Bayesian conditionalisation. Even when you accept $A \rightarrow C$ on independent grounds, upon learning $\neg C$ you may choose either to keep endorsing the inference and conclude $\neg A$ or abandon the inferential principle sustaining $A \rightarrow C$ in the first place. Again, the acceptability of $A \boxrightarrow C$ will go hand in hand with that of the inference and not with the prior probability bestowed upon $A \rightarrow C$.[6]

As always, an example may help make this point clearer. Suppose that a friend of yours is baking a cake and asks for your meaning about three options for the filling—cream, marmalade or chocolate. You believe that unless he goes for chocolate, the cake will wind up quite hopeless. You

[6]Since we have argued that material conditionals can properly replace indicatives in contexts where, as opposed to everyday reasoning, the logical power of premises cannot become impaired by further assumptions (see Section 6.3.3), the foregoing explains the somewhat surprising fact that material conditionals are so good at rendering the function of counterfactuals in Modus Tollens. For Adams, on the other hand, the use of material conditionals in formalising MT was based on 'doubly questionable assumptions'. (Adams, 1975), p.106.

deem marmalade better than cream, though still pretty bad. Given that your friend does not seem inclined to use cream anyway, you say to him:

(123) If you use marmalade, the cake won't be tasty.

Your friend, however, does use cream in the end and to your great surprise, the cake turns out to be delicious. The point is that upon learning that the cake is filled with cream while being tasty, you may not only give up your prior conditional belief *cream* → *bad*, but very likely also the conditional *marmalade* → *bad*. In that case, you will reject

(124) If you had used marmalade, the cake wouldn't have been tasty.

despite having awarded high prior conditional probability to its indicative counterpart. Note that there might have been no way for you of finding out before actually sampling the cake.

Finally, if you find it distracting that apart from learning that the cake is tasty you also learn that it was filled with cream in the example, think of a variant where you are not sure about the actual ingredients in the filling upon taking a bite. The flavour of the cake surprises you, and you come to believe that it would have come out delicious no matter what.

The morale is that it is strange to talk about conditional probabilities prior to learning new information in order to account for counterfactuals in situations where this very information makes you revise the inference that sustained your earlier conditional probability assignment.[7] Both these and the If-I-were-you examples naturally divert our attention from sheer probabilities towards the inferences that hide behind their values, which are often causal in nature.

7.1.2. Dudman's relocation theory

Both the importance of the inference grounding the conditionals and the connections between indicatives and subjunctives we have seen before (the syntax, the future-future and future-past link) inspired V. Dudman to explore the possibility of redrawing the internal frontiers within the realm of natural language conditionals. This attempt by and large failed, but it deserves our attention for its very audaciousness and some useful insights upon which it drew. It is also important to see that the phenomena we have been inspecting in this chapter do not undermine the traditional

[7] This is also the case with the clever counterexample that Adams levels against his own theory. Interestingly, he argues that the most straightforward amendment must take into account causal relations between dispositional states. (Adams, 1975), pp.129-133

distinction between indicatives and subjunctives, so that we can justify an account of meaning that follows these categories.

'Judgements' vs. 'propositions'

Dudman claims that contemporary literature on conditionals has fallen prey to a fundamentally flawed categorisation of conditionals and, as a result, has further addled our understanding of these structures by constructing a lofty theoretical edifice on rather shaky foundations.

His starting point is the observation to the effect that there is a frequent mismatch between the tense of an English sentence, and the time the message conveyed by the sentence intuitively adverts to. He calls such messages *judgements*.[8] Some examples of judgements include

(125) The countess will go home.

or

(126) The countess might have walked home.

The grammatical tense of (125), according to Dudman, is present, although under its standard interpretation it refers to a future event. Likewise, the tense of (126) is past perfect, whereas the event described, if it took place at all, did so in the past simpliciter. Dudman explains this time shift by positing that the tense of these sentences determines the temporal location of the so-called *factual basis* of the judgement. In forming a judgement, we begin with what we know of some factual historical situation and then we let it develop in our imagination in ways we have learned to expect. After inspecting the result, we reach a *verdict*, which determines our choice of modal. On Dudman's account, judgements are tensed verdicts—the tense of a judgement indicates where on a timeline we should look for the facts to be used as inputs in this thought experiment.

Unlike judgements, Dudman's propositions are statements of hard fact and have a definite truth value—there is no imagination involved in their assessment. That is why the tense of a sentence that expresses a proposition tallies with the time the proposition intuitively is about.[9] It follows immediately that there are no hard facts about the future, only forecasts based on present facts and the expected course of events; a result that agrees with many of our intuitions.

[8](Dudman, 1989), p.593

[9](Dudman, 1984), p.144 and (Dudman, 1989), p.594

In this dissertation I have been more generous with the label of 'proposition' than perhaps most contemporary scholars, not hesitating to award it to the content of practically any acceptable assertion. My foremost reason for doing so is that to say that a certain assertive sentence fails to express a proposition usually strikes us as a very radical statement, while the actual theoretical motivation for making the distinction between propositions and non-propositions tends to be rather subtle. In such cases, I think it more appropriate to talk about different kinds of propositions, or about different qualifications (like hedging or conditioning) to which a proposition can be subjected.

I believe this is also the case with Dudman's theory. What he calls 'judgements' are simply certain modalised statements—the fact that these are often made on the strength of an inference, rather than on direct evidence, doesn't, in my view, warrant setting them radically apart from non-modal ones. Again, the grounds on which a proposition is accepted have nothing to do with truth—truth-talk can easily encompass all legitimate assertions, even in cases where there is no straightforward algorithm that could decide for us whether to endorse them or reject them.

And I find it therefore even less justified to pick out as special those modalised statements that are based on a *certain type* of inference, for instance a predictive one. I really don't think that utterances of the following sentences

(127) He may be one of the invited guests.

(128) He may come to the dinner tonight.

should be regarded as belonging to two distinct categories of assertions.

On the positive side, observe that Dudman's theory of 'judgements' bears quite some resemblance to Kratzer's theory of modals—in both cases, there is a modal base, that is, a set of facts against which the proposition in the scope of the modal is evaluated. The choice of the modal depends on the inferential relation between the base and the proposition in hand; the difference between Kratzer and Dudman resides chiefly in the nature of this relation. In Kratzer, this inference is based simply on the relation of entailment, while Dudman postulates a non-formal predictive inference from a factual basis that need not be made completely explicit, as in Kratzer's case.

I believe Dudman has an upper hand here, as Kratzer is forced to codify the propriety of the inference in question in terms of non-accidental generalisations, with all the associated problems (which we have seen

come to the fore especially with counterfactuals).[10] Reducing non-formal inference to explicit entailment is notoriously difficult—it is often more worthwhile to look for another way of characterising it. We shall explore this phenomenon in what follows, but first we have to take a look at what Dudman has to say about conditional clauses.

Dudman's classification of if-sentences

To be sure, not all 'if-sentences' are *conditionals* in Dudman's idiom—this term is reserved only for judgements with embedded if-clauses. Dudman says that these if-clauses encode a *complication* of the message conveyed by the judgement. That means, they introduce an event that impacts on the imagined development of the situation contemplated as a basis, possibly leading to a huge divergence from the original course of history. Since if-clauses do not alter the location of the factual basis of the judgement, they agree in tense with the main clause; however, they refer to a time which lies somewhere in between the temporal location of the basis and the time the judgement as a whole is intuitively about. This explains the presence of a time-shift also in the subordinate clause.[11] To see how an if-clause can complicate the judgements (125) and (126), consider the following examples:

(125if) If she misses the last bus, the countess will walk home.

and

(126if) If she had not had enough money, the countess might have walked home.

For Dudman, non-conditional if-sentences, on the other hand, are real compounds joining two independent sentences, which can express either a judgment or a proposition. They consist of a 'subsidiary if-string' and an independent sentence. Whereas a genuine conditional expresses only one, if complex, message, in the case of non-conditional if-sentences we have two messages linked in a particular way. The independent message is presumptively affirmed on the hypothesis that the dependent message is true. Examples include typical cases of syllogistic reasoning, such as

(129) If some man is mortal, then some mortal is a man.

[10] See Section 5.2.5.

[11] (Dudman, 1989), p.595

but, for instance, also cases in which the if-clause introduces a presupposition necessary for a felicitous utterance of the independent clause:

(130) If the mayor is married, his wife did not accompany him that evening.

Now, it is clear that for Dudman, given the time-shift in 'does-will' conditionals, they fall into the same category as subjunctives, despite having traditionally been considered as indicatives.[12] A further argument he adduces in favour of this classification is our intuition that most 'does-wills' have a 'had-would' counterpart, which only differs from them in tense.[13]

Assessment

To begin with, observe that Dudman can take in his stride most of the difficulties besetting Adams's account that I have touched upon in the previous section. His account does not make use of the distinction between prior and posterior beliefs and he is thus not compelled to constrain the impact of new information to the 'factual base'. He can easily admit that new information can affect both our take on what the facts are and the way in which we imagine they will develop.

Rather than that, my reservations with respect to Dudman's ingenious analysis have to do with the pride of place it awards to syntax. Syntax can often give us a nudge in the right direction, but we should not expect the accidental means a language has developed for communicating particular contents to be more illuminating than the contents themselves. Indeed, the scope of Dudman's analysis seems somewhat constrained to English. Although a range of time-shift phenomena has been reported for many of them, there does not seem to be a general systematic pattern that would allow seamless extrapolation of Dudman's insights to other Indo-European tongues. And to make matters worse, the linguistic evidence available on the employment of 'fake tense' (Iatridou) seems to corroborate, rather than undermine, the traditional classification of conditionals.[14]

It is precisely the contested 'does-will' conditionals that put on a variety of linguistic forms across languages, departing most radically from what should be expected on Dudman's account. For his analysis, it is essential that English expresses future tense by means of a modal, 'will', so that he can argue in favour of a future time-shift in these constructions. However,

[12](Dudman, 1984), p.148. Bennett endorsed this theory of woulds in his (Bennett, 1988).

[13]See page 184.

[14]See (Iatridou, 2000) and (Schulz, 2012).

this is by no means the rule within many European language families. Romance languages, as well as most Slavic languages, are equipped with a proper future tense and do not have to resort to a modal auxiliary. And whereas Romance languages often require a time-shift in future-oriented if-clauses (though not in the main clause)[15], there is no difference in tense between the main and the subordinate clause of a 'does-will' conditional in most Slavic languages. Nevertheless, these languages do mark very clearly the distinction between 'does-will' conditionals and those that have traditionally been considered as subjunctive.

Another shortcoming of Dudman's theory, expressly acknowledged by him, is that it cannot account for the existence of the so-called 'hybrid' cases. Future-to-past conditionals such as

(131) If you feel sick tomorrow, then the pizza wasn't baked properly.

are diagnosed as anomalous, even 'monstrous' in Dudman's own words.[16] On his account, this 'would-be conditional' contains a complication (hence the time shift in the if-clause), without there being a judgement in which it could be embedded. In practice, none the less, these conditionals, predicted to sound natural on the traditional classification, do not pose any difficulty to their users—the other day, I was told by a repairman returning my bicycle to me:

(132) If you don't come back, the repair was successful.

Crucially, 'wills' can do many things analogous to those that can be done by sentences expressing Dudman's propositions. They can for instance be used to perform all kinds of different speech acts apart from predictions (promises, threats or advice). And just as with 'propositions', the scope of these acts can be restricted by means of indicative if-clauses, which can refer to future, present or past events. On the other hand, 'woulds' seem to be used exclusively in assertions, and have no use without a subjunctive antecedent—if this antecedent against all expectations turns out to be true, the would-conditional doesn't seem to commit the speaker to anything.

All in all, I think Dudman is wrong in classifying sentences according to the grounds on which they are properly affirmed. The fact that claims about

[15] An exception is, for instance, Portuguese, which uses a special future subjunctive tense in the subordinate clauses of future-oriented conditionals. This tense also exists in Spanish, but is nowadays only used in a highly formal register proper to solemn legal texts. Latin itself allowed future tense both in protasis and apodosis, see (Allen and Greenough, 1903), p.323.

[16] (Dudman, 1989), p.600

future can only be based upon predictions doesn't affect their status of assertions about the actual course of events. If at the time they advert to the corresponding assertion about the present is true, they are true, otherwise they are false. That's why the whole time-shift issue is a red herring—it is irrelevant for the function of claims about the future whether they are expressed by a modal or by a special grammatical tense. Subjunctives, on the other, talk about alternative scenarios and therefore do not commit the speaker to any indicative statement.[17]

Let us take stock. We have seen that Dudman's recategorisation of conditionals has misfired chiefly for the following reasons:

(I) it is based on a syntactic phenomenon that is peculiar to English

(II) it associates this phenomenon with the grounds on which conditionals are accepted, not their function in discourse

(III) it fails to account for certain perfectly acceptable conditionals

And as I have already announced, in the next section we shall analyse some further problems that Dudman shares with Adams. There it will become clear why I find the debate on whether had-woulds are merely past-tensed does-wills so fruitful for our purposes.

Regardless of the arguments of the next section, though, we have seen that the intuitive distinction between indicatives and subjunctives has withstood Dudman's attacks. This is good, because it entitles us to construe indicatives as conditional assertions, and subjunctives in turn as devices for assessing the tenability of indicatives. In this manner, I shall interpret non-conditional assertions, indicative conditionals and subjunctives as forming layers, each of which depends for its functionality on that of the previous one.

7.1.3. The benefit of the hindsight

Up to this point, we have focused on problems due to peculiar features of either Adams's or Dudman's theory. In this section, however, I shall systematically examine at least the most conspicuous ways in which our posterior assessment of a counterfactual can diverge from a prior assessment of a future-oriented indicative.

Good grounds for asserting an indicative do not always translate into good grounds for asserting a subjunctive at a later point. Examples that

[17] However, they can be used to justify inferences about the actual state of affairs, because their job is to spell out how the world hangs together.

support this point pose difficulties to both Adams's and Dudman's theory; however, if we abandon the strictures of their excessively narrow take on the problem, we can learn a thing or two about our actual use of counterfactuals. In the rest of the chapter, the results gleaned from this discussion will be put into perspective, shored up with theoretical insights hailing from rival theories.

Example 1: no conceivable past indicative

Adams's theory of counterfactuals relies heavily on the notion of a 'hypothetical past assertion of an indicative', while Dudman's posits a past factual base, with respect to which the subjunctive is evaluated in the very same way as a future-oriented indicative would be if this factual base were actual. Nevertheless, with quite a few counterfactuals it is rather difficult to envisage the past circumstances under which the corresponding indicative could have been affirmed:

(133) If I were you, I would consider taking off that Superman costume.

(134) If Bosnian Muslims were bottle-nosed dolphins, the world would not have allowed their slaughter.[18]

It won't do to shrug these examples off as somehow irrelevant.[19] If-I-were-you-counterfactuals are among those most frequently used by ordinary speakers. And the second example, far from fanciful, was put forward in order to drive home a very dramatic point at the height of the Bosnian war.

The upshot for Adams's account is that it is utterly implausible that the above examples inform about any kind of prior conditional probability of the consequent given the antecedent. On the other hand, neither they are based on an alternative *course of events*, as Dudman's theory requires—a counterfactual such as (133) is not based on an imagination of a past factual situation derailed by my suddenly turning into you. Even better, in

(135) $\{x \mid x \notin B\}$ fails to be a set, for its union with B would be the class of all sets.[20]

there is no question of an imagined temporal development within which the absolute complement of B suddenly becomes a set.

[18] Adapted from (Luttwak, 1993).

[19] See (Adams, 1975), p.104. In a similar vein, (Bennett, 2003), p.284-287.

[20] Adapted from (Enderton, 1977), p.27.

It seems that despite the many analogies between subjunctives and past future-oriented indicatives, we must not think that once we get a grip on the latter, we will be able to explain the former by recourse to a simple time-shift. The inferential principles that sustain the examples in this section do not involve a temporal variable; they are not used to justify predictions. I think it is the other way round—conditional predictions must be based on a certain type of inferential apparatus that can in turn usually also support counterfactuals. Arguably, causal schemata constitute a core element of this apparatus, and that's why so many subjunctives are causal in nature. However, as the examples above witness, causality can't be the last word about them.

The examples above show that a valid inferential principle can be applied even to hypotheses that are so silly that they cannot be plausibly entertained as real epistemic possibilities by any reasonable person, and, as a result, hypotheses extremely difficult to construe as viable indicative antecedents. The fundamental question, which I will attempt to answer in the second part of this chapter, is then how such extravagant claims can still be informative for the audience. Before that, however, we should give some thought to a few more mundane examples.

Example 2: unexpected defeaters

Once we have ascertained that time does not have to play a prominent role in counterfactual reasoning, let us now have a closer look at examples where it does, but which nonetheless put the symmetry between indicatives and subjunctives in question. The first case occurs when the benefit of the hindsight provides you with information that could not have been foreseen, but if it had been, it would have effectively defeated the original inference. Think of the following example: you and a friend are about to take a train at 3.00 p.m. and you say

(136) If we leave at 2.30 p.m., we'll get to the station on time.

However, you become embroiled in household chores, miss the train and finally take the one at 5.00 p.m. Contemplating the landscape quickly sliding past your window, you say to your friend with a sigh

(137) If we had left at 2.30 p.m., we would have got to the station on time.

But your friend knows better. As it happens, there was actually a spectacular accident at 2.40 p.m. on the access road to the station, blocking all traffic for an entire hour. If you had left at 2.30 p.m., you would have got stuck in a jam and missed the train anyway.

I think we can safely assume that at least some events relevant for our predictions cannot be foreseen. Even those that can be anticipated *in principle* are often beyond human reach and can only be learned a posteriori. Since you didn't leave at 2.30 p.m., on the theory I have advanced in the previous chapter, your assertion of (136) is void. However, it is clear that the inference on which you based your assertion failed—the fact that you couldn't have foretold the improbable accident and its consequences doesn't let you off the hook. Your inference has simply been trumped by reality.

These considerations provide the leading idea for what I regard as the most promising approach to subjunctives—I believe their basic discursive function consists in sanctioning the propriety of inferential principles that sometimes underlie indicatives. I think this is the crucial insight that inspires Dudman's, Adams's and Edgington's treatment of subjunctives. In my view, it is mostly due to the fact that they all tried to incorporate it within a much too restrictive framework that it didn't come to full fruition. In order to evaluate the inference in question you don't have to restrict yourself to the information available at the moment when it might have been used in a genuine prediction, as both Adams and Dudman would have it. You can of course help yourself to any facts that only become known *a posteriori*.

Let us return to our example. For the very reason that your assertion of (136) is felt to have been unjustified, (137) uttered at a later point will be considered false. Observe that if you somehow had been able to predict the accident, you wouldn't have given much credence to (136) in the first place. Here it is of course critical, both for the indicative and the subjunctive, that the occurrence of the accident should be deemed *independent* from your leaving at 2.30 p.m. That's where causality enters the picture.

I shall take up the issue of causality in the following section; however, I first have to address J. Bennett's criticism of Dudman' relocation thesis, which also seems to put into question an account of subjunctives as devices designed to evaluate indicative conditionals.

Example 3: incomplete scenarios

J. Bennett adopted Dudman's classification of conditionals, though not all the details of Dudman's proposal in (Bennett, 1988). He based such a move chiefly on the time-shift phenomenon and the purported quasi-equivalence between does-wills and had-woulds. Nevertheless, he retracted his adherence to Dudman's classification in his later work, offering a few arguments as to why the traditional way of drawing the line between conditionals

should be regarded as correct.

Bennett points out that there is an entire class of proper utterances of does-will conditionals that do not support the corresponding counterfactuals: they occur when the antecedent eliminates alternative explanations of a piece of evidence, while the consequent follows from the only hypothesis left standing. The following story will make this point clearer:

> We are standing on a balcony and we can see a truck approaching our block of flats. Being rather short-sighted, I cannot make out what kind of truck it is and I utter the following conditional: 'If the driver gets off and starts shouting "Butane!" in a wild fashion, we'll finally be able to take a warm shower today.'
>
> Variation on an example in (Bennett, 2003), p.345-346

The evidence we are presented with is the arrival of a truck, and my antecedent is meant to make plausible the hypothesis that it belongs to the gas delivery company. The prospect of a warm shower follows from such a hypothesis. However, this, according to Bennett, is not enough to support the counterfactual (138) once the lorry passes the building indifferently:

(138) If the driver had got off and started shouting 'Butano!' in a wild fashion, we would finally have been able to take a warm shower that day.

Assuming that the antecedent of my conditional turns out to be false, the best possible explanation is that the truck belonged to the greengrocer instead, or that it belonged to the gas company, but had no gas bottles left for sale.

Bennett argues that if you nevertheless accept my conditional on different grounds, such as if you believe that the shouts of the driver will wake up the gas delivery man living on the first floor of our building and goad him into doing his duty, your basis will support the counterfactual in (138). Since most does-wills can be accepted on different grounds in the same situation, the necessary link between does-wills and had-woulds cannot be sustained.

I find Bennett's analysis of the example impeccable. I would only like to underscore some analogies with the previous one, so that we can glean some insights for our own take on subjunctives. In both cases there is a piece of information missing that, had you been aware of it, would have made you change your mind about the indicative. In the second example, it was the occurrence of the traffic accident and the ensuing jam; in this one,

it is the fact that the truck belongs (say) to the greengrocer. In both cases this crucial information was out of your reach—either because of its unpredictability or your short-sightedness. This of course does not prevent you from making sensible conditional assertions, just as you are often forced to make regular assertions in the absence of all relevant information—the fact that you can hedge them by dint of a conditional clause makes it even easier. However, when the benefit of the hindsight completes the factual panorama, you may see that the inference you drew, though justified in view of what you might reasonably have been expected to believe in the circumstances, was not supported by them.

7.2. Inferential scenarios

It is high time we looked closer at the inferences that we found at the basis of our utterances of counterfactuals. Up to now, I have been deliberately vague about their nature, but that must stop, lest we should render those considerations largely irrelevant. Everybody admits that some kind of inference underlies our use of subjunctives—what is at stake is the best way of characterising them. My purpose in reviewing all the major recent attempts at such characterisation in Chapters 2-4 was to ram home the fact that despite all the ingenuity and strenuous efforts with which they set about it, most of them fail badly at this task. We also saw that the most fruitful approach, the one that made direct use of information about causal networks, felt a little like cheating. Once you are allowed to use such highly structured input, formal semantics becomes a trite exercise in mathematical modelling.[21]

This should give us pause. If we think about an ordinary counterfactual, such as Edgington's

(55) If I had known you were coming, I'd have baked a cake,

we are often at a loss to know the explicit grounds one may have for such an assertion—it does not always seem to fit into a neat causal pattern. However, when we are given a fully specified inferential system, such as one implemented by an axiomatic system or a causal network, accounting for counterfactual reasoning all of a sudden becomes surprisingly easy. We have seen causal networks in action in Section 5.3; for a non-causal example, just recall

[21] See Section 5.3.

(135) $\{x \mid x \notin B\}$ fails to be a set, for its union with B would be the class of all sets.

You have simply entertained a hypothesis, which, as it happens and possibly unbeknownst to you, is excluded by your system (that the absolute complement of a set should also be a set). After that, you have seen that this hypothesis, in view of the inferential principles validated by the system, leads to a conclusion you already know to be false (that is, that the universal class should be a set). This makes counterfactuals naturally fit for use in Modus Tollens arguments.

I will try to exploit this insight (that the use of counterfactuals against the background of a smooth deductive system is unproblematic) in order to account for subjunctives with no obvious basis in a clear inferential framework. I think the most straightforward way of cashing in this intuition is by admitting that our representation of the world is inferentially structured from the very beginning.

7.2.1. The meaning of subjunctives

On various occasions in this chapter I have already hinted at what I deem the most promising approach to the meaning of subjunctives. Just as with indicatives, this account will most likely disappoint those who have come to expect startling new discoveries from the semantic theory. However, I for one wonder why the meaning of vocabulary that we use on a daily basis should have any surprises in store for us. The interesting thing is not what subjunctives *mean*, but rather which of them are *true* and *why*.

The standard way of proceeding in semantics conflates these two questions, because it analyses the meaning of an expression in terms of its contribution to the truth conditions of the sentences in which it appears. However, we have seen that this strategy leaves a lot to be desired, as

(i) it is based on the unjustified assumption that meanings of complex expressions can be reduced to 'less problematic' ones, which also raises the issue of infinite regress

(ii) it invites trivial analyses (often disguised by sophisticated mathematical paraphernalia)

(iii) meanings of all too many expressions cannot be captured in this way.

The alternative I advanced in the previous chapter advocates explaining the meaning of an expression by identifying the function it performs in

discourse. Any natural language is a highly complex system designed to carry out a panoply of very specific tasks, and words can plausibly be construed as conventional tools to achieve these goals. This was, after all, until very recently the standard way in which meaning used to be explained. It was the advent of formal semantics in natural language analysis, with its extrapolation of methods often more suited to the study of formal languages, that changed the picture and our perception of what constitutes the standard procedures.

I argued that the function of indicative conditional clauses consisted in restricting the validity of the speech act performed by the main clause. We also saw that the main clause of would-conditionals could not be used independently, so this explanation could not be straightforwardly extended to them. Subjunctives seemed to be used to make assertions of a peculiar kind—and now we are in a position to see which.

In line with the considerations in the previous sections, subjunctives are used to state the consequences a hypothesis has in the inferential scenario under discussion. These consequences are calculated by applying inferential principles to the hypothesis as well as to any other relevant propositions inferentially independent of it. That's why the use of subjunctives within explicit deductive systems is so clear-cut.[22]

Of course, this doesn't give you any handle on the truth of particular subjunctives, but then, I believe that is not the job of a theory of meaning. Moreover, you still need to plug in the additional input specifying the inferential system in hand. This is easy when it is fully spelled out, but it is much more difficult to view our everyday representations of our environment as inferentially articulated. However, in the next section I will argue that this is exactly what we should do in order to understand the most common uses of subjunctives.

The much sought-after link between indicatives and subjunctives is suggested by the fact that most indicatives are based on inferences. Leaving biscuit conditionals aside for a moment, the truth of a typical conditional $A \rightarrow B$ cannot as a rule be ascertained by direct observation. Hence the need for an inferential cognitive module for making conditional assertions. Subjunctives can in turn be used to make the properties of such modules explicit—without even having to pretend that the speaker is regarding the antecedent as a live possibility and is willing to commit herself to the consequent should it turn out to be true. The difference between an indicative and a corresponding subjunctive comes to the fore when there is uncertainty about what the relevant inferential scenario actually

[22] See example (135) again.

is. Compare (135) to

(139) If the absolute complement of any set is also a set, then something must have gone terribly wrong in my set theory classes.

We have seen that the impact of indicative antecedents on our beliefs is tantamount to that of new information, as we become committed to the consequent should the condition turn out to hold. In the course of such a revision, all kinds of adjustments can take place—we have to recreate a picture of the world coherent with the new data. In the interesting cases, involving an antecedent contrary to our beliefs, the facts to be accounted for do not only include the conditional hypothesis, but also for instance the fact that we falsely believed its negation. This can have dramatic consequences—you may be forced to radically reinterpret what you thought you knew was the case in the light of this new information.

In subjunctive reasoning, on the other hand, you treat the picture of the world you have fashioned as a given and you explore how it hangs together by testing the impact of different hypothetical interventions. When this inferential scenario is *fixed*, acceptability of indicatives and subjunctives converges. This happens typically with future conditionals, as our predictions are entirely based on the inferential scenarios we use for this purpose. A closer look at how they work in practice is in order.

7.2.2. Knowledge of the past

The phenomenon of explanation

To understand how we think about the future, we paradoxically first have to turn our eyes towards the time gone by. Let us begin with a simple question: have you ever wondered why we are interested in the past at all? What happens in the present can still affect us; knowledge of what will happen in the future enables us to make better decisions. However, whatever took place in the past cannot be changed, and can only exert influence upon us through its present and future consequences—arguably, familiarity with the latter should be all that you need to deal with the challenges posed by everyday life. If describing reality is like painting a landscape, we only seem to need a picture of how it looks like at the present moment. So why are we nevertheless often so keen to know how the landscape we're cruising has changed throughout the time?

A first stab at an answer might suggest that to know what preceded the current situation is necessary to understand it properly. I think this is

roughly correct, but it only shifts the focus of our inquiry—we are still left in the dark about what it means to *understand* a situation.

An example may make this clearer. Suppose that upon returning home this evening, you find one of your windows smashed, with a stone lying inside. You are surprised, and you are looking for an explanation. At first, it should of course be a scientific one along Hempel's lines, but that's fairly trivial business—it doesn't take a brilliant mind to imagine the kind of event and the natural laws governing it that may have left in their wake a hole in a brittle glass pane and a stone on your carpet. No, what you really are after is who flung the stone and what their motivation was. And there do not seem to be any hard scientific laws that might do the trick.

The nomological-deductive theory has been found lacking because it treats rigorous scientific explanation as a model for explanation in general. However, alternative proposals have often indulged in the very same error. I therefore think a more promising approach to explanation is to inspect how they work in everyday life and only with that knowledge analyse the peculiar features they acquire in highly specialised contexts.

Let us get back to our broken window. Why is it important to know what exactly happened? The damage has already been done; knowing how it came about (or anything else for that matter) won't undo it. However, a proper explanation will be crucial in determining what you should do next. Was it the nice neighbours' rogue children? Maybe you should drop by and ask them to see about the repair, taking some of their delicious cake with you to make good for all the trouble. Was it the grumpy neighbour who still bears you a grudge for that unfortunate sewage accident? Maybe you should retaliate by slipping a rotten vegetable in his mail box. Was it the work of a paramilitary militia rampaging through the town it has just conquered? Maybe it's time to forget about the window and take to your heels. A good explanation, therefore, restores your grip on the world and provides you with information that is crucial for predicting what the most useful course of action will be in the circumstances.

Not all that happens is in need of explanation. If nobody throws a stone into your window today, you do not need to find out why it has remained unscathed; things are just as they should be. We don't even think of you as having learned anything new upon your arrival home—it is simply so that your expectations have been tacitly confirmed. On the other hand, if a hurricane blasts its way through your town and your windows weather it without any casualty, that will certainly qualify as something to be explained, if only because it defies our expectations. Surprising facts show us that our grip on the world is deficient, that we lack some information that in turn may be relevant for making new predictions.

Either something happened that escaped our attention, or else we did know all the relevant facts, but failed to judge them correctly. In the first case, a good explanation will consist in pointing out the missing fact; in the second, in stating the relation between the facts you already knew and the one that surprised you. And, of course, combinations of both are perhaps the most common. It follows that it is at least partly the needs of the audience that determines what constitutes a good explanation in a given context. You must know *why* someone is surprised by a fact in order to supply her with the information she needs. There is therefore no detailed blueprint that would fit all cases of explanation, as some theories in philosophy of science would have it. Explanation is essentially dialogic in nature.

The above shows that knowledge of past facts is relevant for our expectations and, as a result, for our predictions. The broken window surprises you because you expected it to be sound, and would have predicted it to be so if asked. A creature that bases all its predictions on present facts might be conceptually possible, but such an arrangement would be extremely impractical—it would have to be aware of too many things at the same time and moreover, it would need some outstanding sensory capacities. Two different processes (think, for instance, of extreme joy and grief) often share the same outward appearance, and an ordinary human can tell them apart only by looking at their distal causes.

We are thus constantly using the information about the world at our disposal, past and present, to make defeasible inferences about the future, more often than not without as much as being aware of them. And the bulk of this takes place in the deepest layers of our consciousness—even the crudest sensory input is interpreted against the expectations about how the world should behave, most of which are pretty much hard-wired in our nervous system.

Think of how you last got scared out of your wits by an unsuspecting family member. It may only have been afterwards that you realised you had assumed you were alone at your place—your expectations had adjusted to a particular scenario even without your explicit noticing. And yes, the odds are that the first thing you did after that mad scream was to demand an explanation: 'What are you doing here? Shouldn't you be at work/school/in prison?' You thought you knew what was going on, but the shock taught you otherwise. Now you are aware that you have to catch up on facts—hence the demand for explanation.

For an example of a really unconscious expectation, just think of any sensory illusion—your brain associates a particular affection of the sensory organs with the cause that regularly produces it. All conscious perceptive

input is inevitably primed by our hard-wired expectations about the normal course of things. There is no other way of making sense of crude mechanical impacts of raw matter upon our bodies.[23]

All the foregoing considerations make it extremely implausible to construe these inferences as ultimately logical in nature. We have seen that we often do not have much conscious control over them, which would arguably have to be the case if the conclusions on which we act were logically deduced from a sufficient set of premises. And since this is basically the same kind of inferences that inferior animals also draw (think, for instance, of the expectations that underlie their spectacular yearly migrations), reducing them to logical consequence would necessitate the attribution of logical prowess to shrimps and dung beetles. I don't think this is the way we want to go.

In theory we are of course aware of the extent to which human life is driven by pre-rational impulses and assumptions. In practice, however, philosophy tends time and again to slip back to the picture of human mind as a giant calculator, where all input and output, as well as the intermediary computations, are directly accessible in terms of linguistic premises and rules for processing them. There is no questioning the appeal of such a clean-cut conception, but it simply cannot withstand closer scrutiny. And this approach remains dubious even if advertised as a 'rational reconstruction' of such practice—I can't see any guarantee that this kind of reconstruction must always be possible.

When it comes to expressing the results of these inferences in language, which can be achieved in different ways, the simplistic picture of mind I have sketched above is directly encouraged by our unreflecting acceptance of the truth-conditional approach to meaning. This is because this way of looking at things doesn't distinguish between what assertions convey and what leads us to make them. When you ask whether an endorsement of an inference is justified, people are naturally prone to treat the underlying argument as an enthymeme, enrich it with alleged tacit premises and check for the logical validity of the outcome. However, the actual inferential process, whether or not it is relevant for the meaning of the analysed expression, might have looked quite different—we usually do not have

[23]By the way, these considerations about 'Lorenzian a-priori' (as expounded for instance in (Lorenz, 1978)) allow us to give a surprising twist to the problem of induction—if the world were not regular in nature, no natural adaptation and thus evolution would be possible, and hence we would never be able to state the problem in the first place. No wonder that especially the more ancient parts of our brain rely heavily on such regularities. Now of course, this is not the kind of justification of induction a traditional epistemologist might be after, but then, I doubt we can ever live up to his sky-high standards anyway.

conscious control of all the variables required for making the argument logically valid. To make the reconstructed inference look better, we are thus often compelled to make it rely on non-accidental generalisations together with tacit *ceteris paribus* clauses in their formulation. The problems with this approach are notorious. Our failure at properly fathoming non-accidental generalisations is perhaps the most striking corollary of this unfortunate mindset.[24]

Predictions and counterfactuals

I believe the moral of the failure of most theories of counterfactuals not based on causal networks is that we cannot successfully account for our use of these expressions unless we regard our representations of reality as intrinsically structured. The phenomenon of explanation, as we have analysed it in the previous section, supplies yet another reason for positing such structure. We do not just learn loads of loose facts about the world—we organise them in patterns, schemes, scenarios and scripts.[25] Unexpected facts are surprising because they do not fit in the script we tacitly expected things to follow. That's why we cannot just shrug our shoulders and simply accept these facts for what they're worth—we need to know how to amend the running script in order to come back to grips with the world.

To understand the past means to know both all the relevant facts and the connections between them. Facts are arranged in schemes or scripts. Subsuming the information at our disposal under such a template often enables us to predict future developments and choose the most suitable course of action. And even when such past facts are not directly related to the immediate future, knowing the ins and outs of a given script will give us an edge next time we are in a situation that warrants its use.

How can you bring out the structure of the facts? There is of course explicit causal vocabulary at your disposal, with verbs such as 'trigger', 'engender', 'spur', 'evidence'... Also, many conjunctions ('because', 'wherefore') and adverbial constructions ('as a result', 'in consequence'...) make

[24]Here I take issue especially with Kratzer's theory of modals. I have never found her addressing the question as to how her theory should be interpreted. In particular, I find it very optimistic to assume that you can always spell out the content of conversational backgrounds linguistically, so that you can reduce necessity to logical consequence and possibility to logical compatibility. I believe that the fact that Kratzer's analysis of counterfactuals seems to have foundered on the very matter of generalisations may be due to this problem.

[25]In what follows, I will use these terms more or less interchangeably.

the connections between facts explicit. However, you can often achieve an equivalent effect by sheer choice of words when talking about what happened. By simply uttering

(140) Oswald shot Kennedy,

you have already told half a story: you imply that Oswald fired a weapon launching a projectile that hit Kennedy, who died as a result. Both the existence and the connection of these three events are suggested by your utterance. Examples of vocabulary capable of conveying such causally-laden information abound:

(141) She finally convinced him to take mambo classes.

(142) They cooked a lentil stew for dinner.

(143) He took offence at her implication.

(144) At last she took her revenge for that spiteful comment of his.

There doesn't seem to be a neat distinction between stating loose facts and describing whole chunks of the tapestry they make up, and in view of the suggestions of the previous section, this shouldn't come as a surprise. At no level is the information we acquire about the world wholly free from our expectations about it.

The relation between causally laden reports of the past and counterfactuals was not lost on some historians. For a long time, the received view in historiography (as in philosophy of language) had been that counterfactual reasoning belonged to the realm of idle speculation and should therefore be shunned by a serious historian.[26] However, it eventually dawned on a younger generation of scholars that to eschew any kind of counterfactual considerations would mean to abandon causal vocabulary altogether—and we have just seen that causality lurks in the most unexpected linguistic hideaways. All in all, without being able to say what followed from what, history books would become intellectually stale compilations of random trivia.

To spell out the connections among past facts means to arrange them into a familiar pattern supplied by a script. In (140), Oswald's pulling the trigger, the gun's firing and the sudden stop of Kennedy's bodily functions are subsumed under the script 'shooting someone'. Since it is assumed that the incident followed this script, the counterfactual hypothesis to the effect that Oswald didn't pull the trigger will block the entire chain of events that

[26]See for instance (Tetlock and Belkin, 1996), pp.3-4; (Ferguson, 1999), pp.4-7; (Lebow, 2000), pp.551-557.

occurred in the wake of his actually pulling it. In general, just as in causal networks, several alternative events can in principle occupy a particular slot in a script, which determines the overall consequences given the other things that happened. This is why scripts can be used both to formulate conditional predictions (in the form of indicative conditionals) as well as to sustain counterfactual speculations (in the form of subjunctives).

I believe this is the main point of contact between the two sorts of conditionals. When we expect the future to follow a certain pattern and it really turns out this way, our prior does-will conditional will correspond neatly to a posterior had-would one. And since scripts anchored in our knowledge about the past and the present are our only handle on what is likely to happen next, future-oriented indicatives and subjunctives are readily seen as conveying the same message. Conditional probabilities are therefore not the last word about conditionals—they are awarded *based on* the information stored in scripts (remember that they cannot in general be derived from simple non-conditional probabilities).

In section 7.1.3 we explored some cases in which the usual analogy between does-wills and had-woulds failed. We saw that in Examples 2 and 3, the difference was due to the subject's insufficient grip on the reality, which was corrected by the benefit of the hindsight. I argued that had we had access to the decisive piece of information in each case when making the conditional prediction, it would have coincided with the counterfactual uttered at a later point. These cases, therefore, do not pose a major problem unless you want to defend a very rigid parallelism between future indicatives and counterfactuals, which is certainly not necessitated by the present casual account.

Example 1 is different, though, since there are no plausible indicative counterparts to counterfactuals of the 'if-I-were-you' variety nor to counterfactuals exploring the consequences of a false mathematical statement for the sake of a *reductio*. This shows that we must conceive the function of counterfactuals broader than as consisting exclusively in the assessment of possible past indicatives. In these cases, inferential scenarios are explored for purposes different from their use in predictions or conditional assertions. Though there is obviously no chance of me taking charge of your life, openly entertaining such a possibility can give you a new idea about how to tackle a particular problem, or at least make you feel better about the decision you have already taken. And as regards hypotheses contravening axioms of the set theory, a counterfactual experiment can clearly show you why such an assumption is a non-starter. There is thus no need whatsoever to posit a difference in meaning between typical 'causal' counterfactuals and those that we use to speak about other sorts of inference, even if the

latter are not likely to undergird indicative conditionals.

Sharing scripts

Counterfactual reasoning about actual events is thus no different from that which takes place within the confines of other inferential systems. We ponder a hypothesis, insert it in the inferential scenario in hand and see where it leads us in view of other independent facts and inference rules. When thinking counterfactually about the past, we change one element in the inferential landscape and run the corresponding script, which tells us what we can expect in such circumstances.

Exploring the associated counterfactual possibilities is part and parcel of learning from the past. As we have already noted, it brings out the structure of events. Crucially, scripts underlying these structures are liable to be shared among people. No coordinated animal activity would be conceivable without shared expectations about natural events and other animals' reaction to them. In lower animals, these expectations are shared mainly as an essential part of their instincts. We humans, however, can not only follow our expectations and handle according to them, but we are also endowed with the marvellous capacity of informing our companions about them in advance, of putting them on the table for public scrutiny. And one way of achieving this objective is by means of counterfactuals.

We usually utter counterfactuals in order to enrich the scripts of our interlocutors. Great many scripts we use on a daily basis are learned, not hard-wired in our nervous system and counterfactuals offer a manner of passing this knowledge on people who lack our experience and expertise. Counterfactual utterances about the past may at first sight appear as mere exercises in Monday morning quarterbacking, and some of them may indeed be quite unhelpful. Yet if you realise how often they pop up in all kinds of public discourse, you will soon be wary of dismissing them as futile by-products of flights of fancy.

Often the lesson imparted by a counterfactual is quite straightforward. For instance,

(145) If you had pulled the cat's tail (as was your original intention), it would have attacked you.

(146) If you had stayed out of the fire, your fringed cowboy jeans might still be intact.

are readily construed as warnings against behaviour that is prone to bring unpleasant consequences in its wake. However, we may just as often make

counterfactual assertions about situations that are too complex or peculiar to repeat themselves in their entirety:

(147) If the Dutch central bank had intervened earlier, the bankruptcy of Vie d'Or might have been prevented.

This does not mean that such affirmations do belong to the category of Monday morning quarterbacking. On the contrary, our interest in these cases resembles a professional chess player's study of famous games—most likely, she will never find herself in the exact same situation that she analyses, but knowledge of some of the possible strategies can be of great help in tackling analogous, even if new challenges. Sometimes we can reasonably estimate the degree of analogy between the new situation and a familiar one and try to control for the confounding variables in applying the old script. And, of course, sometimes it will be the best we can do anyway—if such an attempt at extrapolation fails, at least we'll have one more experience to learn from.

Just as it is necessary that we share many of our expectations with our fellow human beings, it is also inevitable that different live experiences and attitudes will give rise to divergent takes on what the future may bring. In our weaker moments, we are all prone to momentarily slip to magical thinking, unable to shake off the suspicion that the whole world is against us:

(148) If I hadn't forgotten my umbrella, it wouldn't be raining now.

Most of us become immediately ashamed of such puerile thoughts, but there are also those who award them much more credence than they deserve. You only need to enter a debate on world politics, though, to see how much people's expectations can differ. These are opinions that make the rounds in certain circles:

(149) If Germany had not recognised Slovenia and Croatia, Bosnian war could have been avoided.

(150) If it hadn't been for Western meddling, Putin would never have invaded Ukraine.

(151) If Cuba hadn't been subjected to the US embargo, it would be a prosperous country now.

Unreasonable though some counterfactuals may be, there is often simply no knock-down argument to decide in favour or against them. After all, they are based on expectations, and these can be pretty subjective. This

doesn't of course mean that you cannot level good arguments against them, for instance by pointing to similar cases where analogous expectations (would) have failed. However, it may well be so that the ultimate value of our scripts can only be ascertained in practice, and not in theory. It is just as well that most home-made political theorists will never be able to put their ideas to work.

It may appear too radical to think that there is often no guaranteed method to decide between the truth and the falsity of a counterfactual. However, from another point of view, it only amounts to stating the obvious. Many past situations are simply too complex for us to apply a straightforward, universally shared script to them. All counterfactual speculation to that effect will therefore necessarily contain an element of chance. Disagreement and uncertainty, even among reasonable and informed people, are the rule—we should not just shrug them off as some sort of anomaly.

7.2.3. Laws

Some may argue, though, that if the cotenability approach to the truth of counterfactuals (perhaps in some sophisticated Kratzerian variant) eventually succeeds, we will have a decision method to separate the wheat from the chaff after all. That may well be so; however, I am quite sceptical about the prospects of this project.

The main reason is that in propping counterfactuals up with laws you do not really succeed in reducing them to something better understood and more amenable to study; you only shift the burden of expressing generality around. If you think about it, why should a prior inquiry into laws be deemed preferable to studying counterfactuals directly? Our long-standing reductive tradition in formal semantics makes such a strategy appear natural, but at this point we know better than to follow its practice blindly.

A possible reply may be that laws are the chief building blocks of scientific theories, which are our best shot at the constitution of the world. With this I cannot but wholeheartedly agree. I also think that it provides a clue of why explaining counterfactuals in terms of laws is misguided.

Our faculty to make predictions and think counterfactually about worldly events greatly precedes, both historically and conceptually, our capacity to formulate sophisticated scientific theories. The latter explores abstract relations between magnitudes in a fairly general manner; the former, despite containing elements of generality, is still firmly anchored to particular situations. They are of course closely related, but I think they differ as to

their function in our thought, as witnessed by Goodman's problem—the alleged law covering the match-striking scenario does not provide a criterion to distinguish background variables from those liable to be altered by the introduction of a counterfactual hypothesis. However, it is intuitively obvious that the status of 'dryness of the match' and 'presence of oxygen' is quite distinct from that of 'not lighting'. This additional information is simply missing in the law, because laws serve slightly different purposes than counterfactuals.

That's also why we have so many counterfactuals that do not seem to rest upon any law worthy of the name. We have already discussed Edgington's cake example[27], but their number is virtually unlimited:

(152) If you had come, we would have had so much fun!

(153) If he cared at all for me, he'd buy me a fur coat.

(154) If her mother had tried to advise her, she wouldn't have listened.

The job of laws is to make explicit the general principles that underwrite our expectations in particular cases. That's why, if available, laws can serve to justify such expectations. However, sometimes the latter are so parochial that it is difficult to come up with a reasonably succinct general rule they might exemplify. That of course doesn't make the predictions based on them less rational—in real life, there is an end to all demands for justification.

A language game is conceivable, where participants can make predictions and use counterfactuals without partaking in law-talk.[28] Should we therefore consider their ability to use and process counterfactuals in any way impaired? Or should we conclude that while they can use counterfactuals competently, they are at a loss for what these really mean? I have my doubts about the former suggestion, and I don't think I can so much as make sense of the latter. On the other hand, it is much more difficult to imagine users of lawlike generalisations devoid of the capacity of counterfactual thinking. That's why explaining counterfactuals in terms of laws is redolent of putting the cart before the horse.

As I've said, I believe that one of the reasons why attempts to reduce counterfactual reasoning to logical inference featuring laws and relevant conditions seem so appealing is our striving for objectivity. We would

[27] See page 55.

[28] Actually, I don't recall my grandparents ever using something like a non-accidental generalisation—in any case, they are very frugal with them. They do use counterfactuals, though, with great success (witness all the wonderful hypothetical consequences of my having grown up in the country).

like to have objective criteria to decide which counterfactuals are true and which are false. We tend to think of laws as completely subject-independent, determined only by the world and discovered by natural scientists—so, the reasoning goes, if we could find a way to distinguish relevant conditions from the variable ones, counterfactuals would be redeemed from all suspicion and awarded the seal of approval of respectable discourse.

Behind this reasoning again hides a view of language that likens it to one particular Alpine country.[29] The attitude of most theorists is 'give me the data and I will tell you which counterfactuals are true and which false.' I am afraid, however, that life, and hence language, are not as simple as the Switzerland conception would have us think. And fortunately so—otherwise they would be hopelessly boring. A more fitting metaphor for language would be a wild forest, whose apparently haphazard structure flexibly adjusts to the terrain and climate conditions, always liable to accommodate new circumstances. And these can vary as much as its diverse inhabitants, who are often compelled to pull off all kinds of improbable stunts in order to get by. Although we are quite adept at finding our way through the familiar spots, we cannot encompass the forest in all its vastness and unpredictability.

To return to our problem, it is not the world, it is other people who evaluate our utterances and judge them true or false. Sometimes they will have good reasons for doing so, sometimes they won't; sometimes they will be certain and sometimes they will opt for an alternative only because suspending belief is not an option in the circumstances. They will accept your counterfactual if they credit you with the experience sufficient for forging reliable expectations; they will endorse it if their expectations tally with yours.[30] Ultimately, the scripts you use for your predictions will be tested in practice—but since you cannot rerun history (let alone implement some impossible antecedents), you might never get to know exactly what would have happened in the one situation over which there is disagreement. This doesn't mean that we cannot rationally argue about our counterfactuals—it only casts a shadow of doubt on the possibility to provide them with definitive grounding.

Laws will not help with this task for one simple reason—the *ceteris paribus* clauses. Sometimes they are construed as magical formulas that can make all the application problems go away. But you should be wary of

[29] See Section 6.1.1.

[30] This, I think, is the point Edgington makes about objective probability in (Edgington, 2007), p.181.

magical panaceas if you live in a wild forest. What a *ceteris paribus* clause really says is 'that there in principle may be relevant variables outside our control, but we are going to ignore them for now'. There is a plethora of conditions that can make the inference from 'presence of oxygen' and 'striking a dry match' to 'the match will light' fail and we will never be in a position to control them all. The assumption that the conditions are normal and the *ceteris paribus* clause is fulfilled, wherefore the law can be applied, *is never risk-free.* Our most reasonable expectations can go wrong, and among them figure those about the world following a particular law-like regularity.

Suppose my grandmother doesn't know that oxygen has to be present for something to catch fire. Is her utterance of

(42) If this match had been struck, it would have lighted.

therefore somehow deficient? I don't think so. She identified the situation as one in which the corresponding expectation would be warranted. I happen to know that the presence of oxygen is necessary for the match's lighting, but I am aware that there was plenty of oxygen around, so my grandmother was right. There may be other defeating conditions I am not aware of, and I may therefore at some point get corrected by someone who knows better, or else a failed prediction may teach me that checking for oxygen and dryness is not always enough. In the meantime, though, both my grandma and I will have to get by on the meagre knowledge we do possess. The point is that I am not substantially better off than her.

If you object that 'a complete description of the world' would settle these questions once and for all, you are already embarked on a flight from Amazonia to Bern. There are no complete descriptions of all the paths you can take through the ticket—surprise is an essential part of a forester's life.

7.2.4. Backtracking, ramps and other red herrings

The problem

There is yet another case of a successful theory of counterfactuals giving rise to spurious problems. We have seen that similarity semantics is to a large extent hollow and apart from the logic (which it was specifically designed to yield), it doesn't make any substantial predictions about concrete counterfactuals.[31] However, even though presented as a mere 'framework for the formulation of the substantive problems'[32], similarity semantics

[31] See Section 3.2.

[32] (Stalnaker, 1984), p.129

based on possible worlds still manages to engender theoretical puzzles that bear no relation to our actual use of counterfactuals.

When evaluating a counterfactual such as

(67) If Oswald hadn't killed Kennedy, no one else would have,

similarity theorists tell you to look at the closest world where Oswald fails to kill Kennedy and 'check' whether there is anyone else that kills him there. This looks pretty abstract and innocent, but we have already seen that there is a tension between the desideratum of closeness and that of supporting the antecedent. A similarity theorist will strive to keep as many actual facts, especially the past ones, true in the selected world(s), and so he will be tempted to come up with a story of how the divergence from actuality comes about. That is where the 'small, localised, simple' miracles of Lewis's make their entrance.[33] In Bennett's idiom, you have to devise a 'ramp'[34] that connects the actual course of history with the counterfactual event specified by the antecedent.

So in the case of (67), you have to think of a last-minute accident that prevents Oswald from killing Kennedy. Either he gets overwhelmed by remorse while aiming, or his hand shakes involuntarily in the crucial moment, or a gust of wind slightly diverts the trajectory of the bullets. The most plausible of these stories is supposed to win the contest for 'closest'.

But why stop there in trying to find out what would have to have been different for the antecedent to come out true? Why should strange last-minute accidents have primacy over more profound structural differences that nevertheless do not require miracles? Maybe if Oswald hadn't killed Kennedy, it would have been because he had been mentally stable in the first place. P. B. Downing suggested a more plausible example, paraphrased by Lewis:

> Jim and Jack quarrelled yesterday, and Jack is still hopping mad. We conclude that if Jim asked Jack for help today, Jack would not help him. But wait: Jim is a prideful fellow. He never would ask for help after such a quarrel; if Jim were to ask Jack for help today, there would have to have been no quarrel yesterday. In that case, Jack would be his usual generous self. So if Jim asked Jack for help today, Jack would help him after all.
>
> (Lewis, 1979), p.456. Example from (Downing, 1959).

[33] (Lewis, 1979), p.472
[34] (Bennett, 2003), p.214

Bennett coined the name 'backtracking arguments' for considerations akin to the above one.[35] Their common trait is that they change the past in order to accommodate a counterfactual hypothesis. As Downing's puzzle shows, such a change can in turn lead to the endorsement of a counterfactual that we had confidently rejected before engaging in 'backtracking'.

Lewis's explains this phenomenon by admitting a non-standard vagueness resolution of the similarity relation that is triggered in 'some special contexts', for instance, if someone puts forward such a counterfactual. In a similar vein, A. Arregui analyses most backtrackers as featuring layered modalities, as they tend to sound much more acceptable if the 'would' in the consequent is complemented with 'have to'.[36] According to Arregui, in backtrackers the original standard similarity relation is composed with the accessibility relation corresponding to an ontic necessity modal $have_{mod}$.[37] The embedded modality's task is to 'shift the emphasis back to laws'[38], since Lewis's standard similarity relation doesn't frown upon small miracles. Bennett, on the other hand, came eventually to regard backtrackers as nothing out of the ordinary—backtracking and ramps amounted for him essentially to the same phenomenon.[39]

Although there are undoubtedly valid intuitions at work in these explanations, I find them unsatisfactory for the very same reason I have been hammering on throughout this essay. It may sometimes be worthwhile to draw up a neat formal model of the semantic contribution of a piece of vocabulary, but it seldom provides deep insight into why it is actually used. The same goes for backtracking—we are left in the dark about why the standard similarity resolution with 'tiny miracles' is so strongly preferred by normal people to the law-abiding backtracking one. I will try to throw some light upon this question in what is left of this section.

Lewis's example gives us a hint. If I ask you 'If Jim asked Jack for

[35](Bennett, 1973), p.391

[36]Actually, if you take a close look at Lewis's rendering of Downing's example, you will find that at no point in the story a mere 'would'-counterfactual about the past is asserted.

[37]See (Arregui, 2005b), Chapter 3.

[38](Arregui, 2005a), p.13.

[39](Bennett, 2003), pp.276-287. This came after a detour through a completely symmetrical theory with respect to the past and the future in (Bennett, 1984). I will henceforth disregard Bennett's final (for the time being) theory of backtrackers and assume that they do require a 'different resolution of similarity' than regular counterfactuals. Bennett's theory predicts with respect of Downing's puzzle that assertions of 'If Jim asked Jack for help today, Jack wouldn't help him' and 'If Jim asked Jack for help today, there would have to have been no quarrel' should be perceived as incompatible. I don't share this intuition.

help today, would Jack help him?', you are obliged to take into account all elements of the current situation in order to reach a verdict. By no means can you leave out the fact that there was a quarrel yesterday. It is irrelevant that Jim would never ask for a favour after such a quarrel—that Jack would not help him might be *Jim's very reason* for refraining from doing so. We always evaluate counterfactuals against the backdrop of a particular situation, even if we are ignorant of some of its details. Past events form part of this backdrop—you cannot change them at will. The quarrel makes Jim's asking for help highly improbable, but it does not rule it out completely, so the counterfactual scenario is easily imaginable.

The fact that our representations of the world come in structured by default exonerates us from racking our brains as to how to accommodate counterfactual hypotheses. Facts are modular from the very beginning (this is belied by their construal as members of the power set of the universe of possible worlds)—we can, as it were, throw some out and substitute them for others without affecting the background. That's why counterfactual backtracking is so controversial and also why there is absolutely no need for ramps, let alone any kind of miracles in our theory. If they are not possible for examples from Section 7.1.3, they are not needed for more common counterfactuals either.

Rationale for ramps

We have seen that this confusion is to a large extent theory-driven, but some of it has intuitive roots as well. Take ramps first. It may admittedly be disconcerting for a person to reflect for the first time on her unstudied use of counterfactuals and see the range of antecedents she has up to now used without second thoughts. 'If my husband were blond, if the Soviet Union had been a democratic country, if Challenger hadn't crashed—how is all that supposed to happen?'. These are traits, states and events that in real life cannot just emerge out of the blue. The temptation to come up with some story to justify use of counterfactual antecedents is therefore considerable, and if on top of that you employ a semantic theory that entreats you to evaluate consequents in alternative world-histories, it will become almost irresistible. However, after understanding a little better why people use counterfactuals in the first place as well as how they work, we know better than that. Real use of counterfactuals doesn't require connecting antecedents with the actual history in any way.

The real reason why sometimes we do engage in supporting our counterfactuals with ramps is not so much to underscore their plausibility as to bring out the *contextual relevance* of our assertion. In view of the role

counterfactuals play in our intellectual life, it is often (though not always) important for the antecedent to have been a live possibility, that is, its negation must not be regarded as overdetermined by the circumstances. If you are to draw a lesson from the past, it usually doesn't help if the alternative explored by a counterfactual was not *really* an option.

Suppose you and your friend are hopelessly lost in a maze at a theme park. It is getting dark, children with lollipops have all left, the park is closing and everybody seems to have forgotten about you. Between sobs, your friend says

(155) If we had a map, we could find the way out.

Talk about irritating! And certainly not because what your friend said is false—on the contrary, its very obviousness makes it a stupid thing to say. The management of the park didn't reckon with the possibility that anyone could be so thick as to get lost in the simple maze, so there are no maps available. Your friend is right, but his contribution is rather inadequate, as it doesn't teach you anything worthwhile about the past.

But let us change the example a little. Imagine now that at the entrance of the maze, there had been girls handing maps to each visitor—you wanted to make an impression and smugly refused to take one. Your friend's assertion is now presented in a new light. It may be just as annoying, but for a different reason. You could easily have got hold of a map—your friend's assertion makes it clear just how bad your decision at the entrance was.

Popular wisdom has long warned that counterfactual reasoning with irrelevant antecedents can easily slide into daydreaming. The point of the many adages concerning unrealisable possibilities, perhaps the funniest example of which is the Polish

(156) If granny had a moustache, she would be the grandpa.[40]

is not the futility of *all* counterfactuals, but only those that explore alternatives that are neither here nor there. Unfortunately, similar examples have bedevilled philosophical literature (think, say, of Bizet and Verdi, or Caesar in the Korean war), upsetting our intuitions and giving rise to distracting puzzles.

Imagined ramps put forward to make a case for a counterfactual's relevance usually start at a decision point or, in general, at a point considered

[40] 'Gdyby babcia miała wasy, toby była dziadkiem.' I learned about this proverb through F. Veltman, who collects them. Two rough English equivalents are 'If "ifs" and "ans" were pots and pans, there would be no work for tinker's hands.' or perhaps the better known 'If wishes were horses, beggars would ride.'

underdetermined by the script used to interpret and organise the events. In this way, the antecedent is more or less guaranteed to refer to a relevant alternative. Regardless of the question of determinism, it is a platitude that we cannot predict everything that will actually happen.[41] A script cannot therefore always reduce the range of possible outcomes to one and has to allow for more of them. The other possibilities, in a sense, remain live even after one is realised. It is through their use in describing reality organised in scripts that ontic modalities ('it could/might have been so') connect with epistemic ones ('it can/may be so').

Possible-world semantics, coupled with the intuition that ramps are sometimes relevant for assessing counterfactual assertions, led Bennett and Lewis to think that each and every worthwhile counterfactual needed such a grounding. And, indeed, for most counterfactuals you can come up with some kind of story that shows that the antecedent was a possibility after all (although in a way irrelevant for most practical purposes). That's the job of the 'small, localised, simple miracles'. 'A particle swerving a little or a neuron firing'[42] are events that typically strike us as unpredictable and thus can serve as universal ramp-initiators. Yet, of course, people had used counterfactuals long before learning about neurons or quantum mechanics.

The use of reasonable antecedents is especially important in counterfactual history. You can certainly speculate about how the existence of a functional telephone network connecting Aztec villages might have affected the course of the Spanish conquest, but you will most likely be found to be playing a child's game. Serious historians are usually more interested in close calls and near misses. In this manner, they can underscore the importance of certain decisions or unexpected events. It also enables them to explore alternatives that historical actors themselves were compelled to (or at least should have) entertain, thus throwing light upon the rationale (or lack of it) behind their decisions.

No wonder that this desideratum features prominently in theoretical work on counterfactual reasoning in history. However, in these writings as well, a counterfactual's germaneness is often confounded with its plausibility. And such a theoretical attitude also appears to percolate among practising historians, as the following passage in Tony Judt's *Postwar* attests:

[41] Actually, determinism may be construed as the thesis that there is a gigantic global script that encompasses all our partial ones and regulates all events. Seen in this light, it dangerously resembles wishful thinking, a forest prowler's daydream of a tranquil life in a Swiss village.

[42] (Bennett, 2003), p.217

> The appeal of neutrality-as an alternative to defenceless confrontation-was growing, in Germany and France alike. If the Korean War had not broken out just at this moment (a reasonable counter-factual, since it nearly didn't) the contours of recent European history might look very different indeed.
>
> (Judt, 2005), p.151.

Here, 'reasonable' must of course be read as 'relevant'. Korean war was (or was not) of paramount importance in shaping recent European history regardless of whether it was a close call or not.

Rationale for backtracking

Let us now turn to backtracking. I have defended, *pace* Bennett, that it is a slightly different phenomenon from ramps. A ramp suggests a way of changing the past in order to show that the antecedent of an asserted subjunctive 'was something to be reckoned with in the circumstances'; backtracking changes the past in order to make it possible for the antecedent to come true.

Let us recall Downing's puzzle (page 216). Given the tiff, Jack wouldn't help Jim if asked to do so. Of course, Jim most likely wouldn't ask Jack in the first place, but it is only up to him—he could change his mind at any moment if he wanted to. However, he is aware that such a change of mind would be fruitless—that's why for him to ask, there would have to have been no quarrel yesterday or they would have to make up first.

Suppose now that they still haven't made up. You may find yourself inclined to give different answers to

(157) If Jim asked Jack for help today, would Jack help him?

and

(158) If Jim asks Jack for help today, will Jack help him?

By most people lights, as the literature shows, the counterfactual is false. However, you can argue a case for the indicative being true. The reason is that when making a conditional prediction, you have to take into account the fact that for Jim to ask Jack for a favour, it is necessary that they should have made up first. On the other hand, in order to evaluate the counterfactual, you need a complete inferential scenario—and since it is not at all certain that Jim and Jack will make up, you may stick to what things are like at the present moment. Of course, the indicative will be

forced to take the same value as the counterfactual if we replace 'today' with 'now'.[43]

This example suggests that to ask what would/will have to happen for Jim to ask Jack for help or for Jack to help Jim is to make a slightly different kind of question than the above counterfactual poses—one about the conditions required for something to happen. As we have seen, these can be relevant in evaluating indicatives. They are also vital for planning, as they point out the course of action that must be pursued if a particular goal is to be achieved. Just as the standard counterfactual reasoning, the search for necessary conditions is based on our expectations and hence on scripts. This is further witnessed by the fact that they support counterfactuals of the form 'the goal would not (have) be(en) achieved if the necessary conditions weren't fulfilled'. However, while regular counterfactuals explore the consequences a script associates in a given situation with the event introduced by the antecedent, no matter how it comes about, the search for necessary conditions looks for events that must occur for the contemplated outcome to take place. The former looks forward, the latter backward.[44]

This difference looms large also on the grammatical level. As we have seen, it has often been remarked that in general, backtrackers sound better with a modalised consequent. Thus,

(159) If Jim asked Jack for help today, there would have been no quarrel yesterday.

is much better rendered as

(160) If Jim asked Jack for help today, there would *have to* have been no quarrel yesterday.

While this is certainly an improvement, I still think it stops halfway. The really adequate rendering of the intended proposition is, in my view, as follows (this is also the syntactic structure I have inconspicuously been using the whole time):

(161) *(In order) for* Jim *to* ask Jack for help today, there would *have to* have been no quarrel yesterday.

[43] An analogous point seems to be made in (Woods, 1997), p.86, but I think Woods is wrong about the indicative conditional, which he assumes to be on the same footing as the subjunctive at all times (he follows Dudman in setting does-wills apart from the rest of indicatives). None the less, Woods does suggest that the result of the Ramsey-test would in this case be at odds with the procedure for evaluating does-wills and had-woulds.

[44] The meaning of 'necessary' and 'sufficient condition' in colloquial language tends to be time-bound. That's why the logician's abstract use of these notions sometimes engenders perplexity in non-sophisticated users.

So in the most favourable circumstances, backtrackers' difference from standard counterfactuals is brought out in the dependent clause as well as in the main one. This is a phenomenon encountered across a number of languages (for simplicity I only use forward-looking examples):

(162) *For* Jack *to* help Jim, they would have to make up first.

(163) *Para que* Juan le ayudase a Jaime, tendrían que reconciliarse antes.

(164) *(Nato) aby* Ján pomohol Jakubovi, museli by sa najprv pomeriť.

(165) *Opdat* Jack Jim zou helpen, moesten zij eerst tot een verzoening komen.

My suggestion is, therefore, that bare backtrackers can sometimes squeak through as sloppy ways of expressing necessary conditions for the eventuality stated in their antecedents. This is enabled by the fact that the proper constructions for conveying such messages are close relatives of counterfactuals, coupled with the fact that counterfactuals can in general be evaluated against different inferential scenarios (not only rough causal scripts). Just because standard causal resolution is the default setting for counterfactuals about actual events, in order to avoid confusion, language provides special syntactic structures to express backtracking inferences.

7.2.5. Fatalism

To close this chapter, and with it the whole essay, I will try to show that what we have learned about indicatives and subjunctives can give us insight into what is wrong about the famous 'fatalist argument', as exposed by Dummett:

> The standard form of the fatalist argument was very popular in London during the bombing. The siren sounds, and I set off for the air-raid shelter in order to avoid being killed by a bomb. The fatalist argues, 'Either you are going to be killed by a bomb or you are not going to be. If you are, then any precautions you take will be ineffective. If you are not, all precautions you take are superfluous. Therefore it is pointless to take precautions.'
>
> (Dummett, 1964), p.344

Of course, nobody has ever been swayed by this reasoning, but nevertheless, scholars have found it surprisingly difficult to agree on which of the above inference moves should be branded as invalid. Since the argument essentially relies on conditionals, this is another indication of how

little the profuse literature on the subject has furthered our understanding of them.

As with any paradox, our task is twofold—on the one hand, we must identify the source of the problem and on the other, it behoves us to explain the intuitive appeal of the invalid reasoning. I shall begin with a schematic representation of the argument. Let us define K as 'You are going to be killed by a bomb', P as 'You will take precautions'[45], I as 'Any precautions will be inefficient' and S as 'Any precautions will be superfluous'. Then the argument goes as follows:

1. $K \vee \neg K$	Tautology
2. $K \rightarrow I$	Premise 1
3. $\neg K \rightarrow S$	Premise 2
4. $K \rightarrow (I \vee S)$	WC; 2
5. $\neg K \rightarrow (I \vee S)$	WC; 3
6. $I \vee S$	$E^{\vee}$; 1,4,5

Dummett himself believed that Premise 1 and Premise 2 must feature different types of conditionals. Stalnaker put the problem down to the rule of disjunction elimination, which allegedly merges together two indicative conditionals that are both reasonable, but in slightly different contexts. I will try to show that, fortunately, the solution to the puzzle is much more simple and does not require sophisticated tinkering with our logic nor the hackneyed recourse to polysemy.

I actually think that it is quite obvious that the main culprit in the above derivation is Premise 2. It simply is not true that if you survive the raid, the precautions will have been superfluous. You may well survive only thanks to having taken them. On the other hand, Premise 1 seems OK—if you do not survive, then any precautions you may have taken must have been inefficient indeed.

Now for the second, more interesting task—we have to elucidate why Premise 2, despite being false, has been able to baffle such eminent minds for so much time. To pin down the source of its apparent plausibility, let us reconstruct the reasoning that may lead to its acceptance. We will proceed by conditional proof, but we will avoid importing outside premises that renders it probabilistically invalid.[46]

[45] This proposition will not appear in the reconstruction itself, but will be useful in the subsequent analysis.

[46] For a long time I had thought that the fallacy was due to an inconspicuous application of a probabilistically invalid inference rule to complex conditionals. I now think that the problem with the derivation is much simpler, and is not logical in nature.

So assume for a moment that you will not be killed in the raid. Given this assumption, whether or not you take precautions, you will survive the raid. Well, the reasoning goes, this is exactly what it takes for the precautions to be superfluous, isn't it? Ergo, assuming that you will not be killed in the raid, the precautions will be superfluous.

To see what is wrong with this argument, just transpose it into the past tense. Suppose the raid is over and you are looking for a friend. Either he was killed in the raid or he survived. If he got killed, then the precautions he may have taken must have failed. But if he survived, then he did so even if he hadn't taken any precautions, no matter how long the odds. Does this mean that if he survived, the precautions he might have taken were superfluous? Suddenly it doesn't strike one as plausible at all. Whatever happened?

What we really need in the past scenario in order to conclude that the precautions were superfluous is not an indicative, but rather a subjunctive:

(166) If he hadn't taken any precautions, he would still have survived the raid.

This is because the subjunctive explores the general structure of the raid script, unfettered by how it actually ran. And 'superfluous' characterises an action with respect to a script as a whole—it states that the relevant outcome was or would have been unaffected by it.

Usually we do not need scripts in order to obtain information about the past. Someone may simply tell you that your friend survived, or you may make him out in a crowded street. None of this will make you conclude that the precautions he most likely took to stay alive were superfluous. However, the situation is different with respect to our knowledge of the future, which typically *is* based on scripts. Therefore, when we are told that 'we will survive even if we do not take precautions', our tendency is to interpret this information as being sustained by the relevant script and thus warranting the conclusion that the precautions are really superfluous. Since this is the paradigmatic scenario, future indicatives and corresponding subjunctives usually share their truth value. However, in our case the indicative is not based on a script, but only on a prior hypothetical assumption—hence the confusion.[47] The corresponding subjunctive is unaffected by the assumption and may well be false.

To sum up, the indicative $\neg P \rightarrow \neg K$ doesn't in general warrant belief in S, just as its past variant doesn't warrant belief in the past variant of S.

[47] This, by the way, is not the only possibility—the same goes for a case in which a time traveller tells you that you will survive. That's why this difficulty is not due to an indiscriminate use of conditional proof.

We need a subjunctive to do the job. In this special case, the truth value of the future subjunctive may diverge from that of the indicative. We are led to confusion because this is so exceptional in future conditionals. In past conditionals, on the other hand, the divergence is common, and therefore the past version of the argument fools no one.

Conclusion

If you have made it thus far, I'd better not keep you much longer. Let me just sketch with broad strokes the central theses of the last two chapters:

1. The use of model-theoretical methodology in the study of natural language invites the idea that the expression(s) under study must be paraphrasable in a more elementary vocabulary.

2. Since nothing guarantees the existence of such a vocabulary, many analyses surreptitiously smuggle in ingredients dependent for their interpretation on the analysed expression.

3. Truth-conditional approaches generally conflate the meaning of an utterance with considerations about its justification. They often tacitly assume that translations into the privileged vocabulary can be straightforwardly checked for their truth.

4. Rather than that, I suggest that the meaning of an expression is more naturally explained in terms of the standard function it performs in a conversational exchange. This holds not only for non-truth-conditional vocabulary, but also for a plethora of truth-conditional expressions.

5. In view of the above, indicative conditional clauses are best viewed as devices for suspending the performance of the speech act conveyed by the main clause, regardless of its illocutionary force.

6. Subjunctive conditionals, on the other hand, are used to make simple assertions about the inferential potential of rejected propositions against the backdrop of a contextually salient inferential scenario.

7. Our takes on what the world is like are at all times inferentially articulated—they guide our expectations and predictions. Information about this inferential articulation is stored in *scripts*. Through them, subjunctives connect with future indicatives.

This thesis was born out of frustration. My chief goal in writing it was to explain, or perhaps I should say rationalise, the lack of enthusiasm with which I assimilated most of the current literature on conditionals, despite finding the topic to be of an extraordinary importance. I do think the course that has been set must change dramatically; however, to borrow an aphorism, it is always the rearguard that complains that we should be

running in the opposite direction. So maybe it's just that I have failed to appreciate the value of the work in this field, lacking the ability to contribute in any worthwhile manner. It is for you to decide.

Appendix A

Some technicalities for Chapters 4 and 5

A.1. Stalnaker

$\preceq_\alpha$ is a well-order

We will check for antisymmetry, strong connectedness, transitivity and well-foundedness.

a) *Antisymmetry.* Assume $\beta \preceq_\alpha \gamma$ and $\gamma \preceq_\alpha \beta$. Then, by definition, $f(\{\beta, \gamma\}, \alpha) = \beta$ and $f(\{\gamma, \beta\}, \alpha) = \gamma$. By definition of a set by extension and since f is a function, $\beta = \gamma$.

b) *Strong connectedness.* Take any β and γ. By Constraint 1, $f(\{\beta, \gamma\}, \alpha) \in \{\beta, \gamma\}$. Hence, $f(\{\beta, \gamma\}, \alpha) = \beta$ or $f(\{\beta, \gamma\}, \alpha) = \gamma$. In the first case, $\beta \preceq_\alpha \gamma$, in the second, $\gamma \preceq_\alpha \beta$.

c) *Transitivity.* Assume $\beta \preceq_\alpha \gamma$ and $\gamma \preceq_\alpha \delta$ and suppose $\beta \not\preceq_\alpha \delta$ for reductio. By strong connectedness, either $\beta \preceq_\alpha \delta$ or $\delta \preceq_\alpha \beta$, so by disjunctive syllogism $\delta \preceq_\alpha \beta$. I will show that this can only occur if $\beta = \gamma = \delta$. By Constraint 1, $f(\{\beta, \gamma, \delta\}, \alpha)$ equals either β, γ or δ. Suppose it's β. As $\beta \in \{\beta, \delta\}$, $f(\{\beta, \delta\}, \alpha) = \delta$ and $\delta \in \{\beta, \gamma, \delta\}$, by Constraint 4, $\beta = \delta$. But we know that $\gamma = f(\{\gamma, \delta\}, \alpha) = f(\{\gamma, \beta\}, \alpha) = \beta$. So $\beta = \gamma = \delta$. The argument for $f(\{\beta, \gamma\, \delta\}, \alpha)$ equal to γ or δ is analogous. By strong connectedness, we have $\beta \preceq_\alpha \beta$, and together with the above equality, $\beta \preceq_\alpha \delta$ after all.

d) *Well-foundedness.* We shall show that for any A such that $[\![A]\!]_\alpha \neq \emptyset$, $f(A, \alpha)$ is the unique least element of $[\![A]\!]$. Set $\beta = f(A, \alpha)$. By Constraint 1, we know that $\beta \in [\![A]\!]$. Suppose that there is a $\gamma \in [\![A]\!]_\alpha$ such that

$\gamma \preceq_\alpha \beta$. Hence, by definition of $\preceq_\alpha$, $f(\{\beta, \gamma\}, \alpha) = \gamma$. But then we have both $f(A, \alpha) = \beta \in \{\beta, \gamma\}$ and $f(\{\beta, \gamma\}, \alpha) = \gamma \in [\![A]\!]$, so by Constraint 4, $\gamma = \beta$ and β is the unique least element of $[\![A]\!]$. This also warrants the assertion that f always selects the closest antecedent world to the actual world.

⊣

Proofs of the invalidity of certain inference patterns on Stalnaker frames

Antecedent strengthening: $A > C \nvdash (A \wedge B) > C$
Let α, β, γ be the only worlds in K and let $\alpha \prec_\alpha \beta \prec_\alpha \gamma$.[1] Set $\beta \in [\![A \wedge \neg B \wedge C]\!]$ and $\gamma \in [\![A \wedge B \wedge \neg C]\!]$. Then you have that $A > C$ is true, while $(A \wedge B) > C$ is false at α.

Transitivity: $A > B, B > C \nvdash A > C$
Let K be as before. Set $\alpha \in [\![\neg A \wedge \neg B \wedge \neg C]\!]$; $\beta \in [\![\neg A \wedge B \wedge C]\!]$; $\gamma \in [\![A \wedge B \wedge \neg C]\!]$. Then you have that $A > B$ and $B > C$ are true, while $A > C$ is false at α.

Contraposition: $A > C \nvdash \neg C > \neg A$
Take again K as before. Set $\alpha \in [\![\neg A \wedge C]\!]$; $\beta \in [\![A \wedge C]\!]$; and finally $\gamma \in [\![A \wedge \neg C]\!]$. At α, you have $A >$ (still) C together with $\neg C > A$.

Simplification of Disjunctive Antecedents: $(A \vee B) > C \nvdash (A > C) \wedge (B > C)$
Let K be as always and set $\alpha \in [\![\neg A \wedge \neg B \wedge \neg C]\!]$; $\beta \in [\![A \wedge \neg B \wedge C]\!]$; $\gamma \in [\![\neg A \wedge B \wedge \neg C]\!]$. As a result, $A > C$ as well as $(A \vee B) > C$ are true, while $B > C$ is false in α.

⊣

Antecedent strengthening makes all conditionals strict

$A > C \equiv \Box(A \supset C)$

Sufficiency

1. $A > C$	Premise
2. $(A \wedge \neg C) > C$	From 1 by AS
3. $(A \wedge \neg C) > (\neg A \vee C)$	From 2 by CW
4. $\neg(A \supset C) > (A \supset C)$	From 3 by def $\supset$ and REa)b)
5. $\Box(A \supset C)$	From 4 by def $\Box$

Necessity

[1] Here, $\beta \prec_\alpha \gamma$ is of course defined as $\beta \preceq_\alpha \gamma \wedge \beta \neq \gamma$.

1. $\neg(A \supset C) > (A \supset C)$	Premise
2. $(A \wedge \neg C) > (A \supset C)$	From 1 by REa) and def $\supset$
3. $(A \wedge \neg C) > A$	CI + CW
4. $(A \wedge \neg C) > ((A \supset C) \wedge A)$	From 2,3 by CC
5. $(A \wedge \neg C) > C$	From 4 by REb)
6. $(A \wedge C) > C$	CI + CW
7. $((A \wedge C) \vee (A \wedge \neg C)) > C$	From 5,6 by AD
8. $A > C$	From 7 by REa)

⊣

A.2. Lewis

Partial truth conditions

TC Counterfactuals No Uniqueness Assumption I
$[\![A \ \Box\!\!\rightarrow C]\!]^{\mathcal{M}} = \{\alpha \in K \mid [\![A]\!]_\alpha = \emptyset \vee (\forall \beta \in [\![A]\!]_\alpha)((\nexists \gamma \in [\![A]\!])(\gamma \prec_\alpha \beta) \supset \beta \in [\![C]\!])\}$

TC Would-Counterfactuals No Uniqueness nor Limit Assumption
$[\![A \ \Box\!\!\rightarrow C]\!]^{\mathcal{M}} = \{\alpha \in K \mid [\![A]\!]_\alpha = \emptyset \vee (\exists \beta \in [\![A]\!]_\alpha)\forall \gamma(\gamma \preceq_\alpha \beta \supset \gamma \in [\![A \supset C]\!])\}$

TC Might-Counterfactuals No Uniqueness nor Limit Assumption
$[\![A \ \Diamond\!\!\rightarrow C]\!]^{M} = \{\alpha \in K \mid [\![A]\!]_\alpha \neq \emptyset \wedge (\forall \beta \in [\![A]\!]_\alpha)\exists \gamma(\gamma \preceq_\alpha \beta \wedge \gamma \in [\![A \wedge C]\!])\}$

General truth conditions

The truth conditions given for the case without the Uniqueness and the Limit Assumption prescinded from the existence of genuine incomparabilities, and thus would in their presence change the truth value of some counterfactuals with respect to the truth conditions given before. Intuitively, think of several incomparable chains of A-worlds. Pursuant to the truth conditions we employed prior to giving up the limit assumption, $A \ \Box\!\!\rightarrow C$ would come out true only if C were true in the closest world of each of the chains. However, now we would only have to find one chain with a world k below which all words are C-worlds, and forget about the other chains. This is definitely not what we want. A simple tweak would involve requiring instead that any chain whatsoever should contain some such k—this would be more in line with the intuition underlying the original truth conditions. The following formulation implements this insight:[2]

[2]See (Lewis, 1981), p.230.

TC Would-Counterfactuals General
$[\![A \mathrel{\Box\!\!\rightarrow} C]\!]^{\mathcal{M}} = \{\alpha \in K \mid [\![A]\!]_\alpha = \emptyset \vee (\forall \delta \in [\![A]\!]_\alpha)(\exists \beta \in [\![A]\!]_\alpha)(\beta \preceq_\alpha \delta \wedge \forall \gamma(\gamma \preceq_\alpha \beta \supset \gamma \in [\![A \supset C]\!]))\}$

and for might-counterfactuals

TC Might-Counterfactuals General
$[\![A \mathrel{\Diamond\!\!\rightarrow} C]\!]^{\mathcal{M}} = \{\alpha \in K \mid [\![A]\!]_\alpha \neq \emptyset \wedge (\exists \delta \in [\![A]\!]_\alpha)(\forall \beta((\beta \in [\![A]\!]_\alpha \wedge \beta \preceq_\alpha \delta) \supset \exists \gamma(\gamma \preceq_\alpha \beta \wedge \gamma \in [\![A \wedge C]\!]))\}$

On the logical level, admission of incomparabilities can be accomplished by pruning away *Strengthening with a Possibility*:

(ASP) $(\neg(A \mathrel{\Box\!\!\rightarrow} \neg B) \wedge (A \mathrel{\Box\!\!\rightarrow} C)) \supset ((A \wedge B) \mathrel{\Box\!\!\rightarrow} C)$

or, more simply, using might-counterfactuals:

(ASP$^{\Diamond}$) $((A \mathrel{\Diamond\!\!\rightarrow} B) \wedge (A \mathrel{\Box\!\!\rightarrow} C)) \supset ((A \wedge B) \mathrel{\Box\!\!\rightarrow} C)$

Is there a case for rejecting ASP? There is a neat counterexample in (Ginsberg, 1986), p.50. The following argument doesn't sound valid:

> If Bizet and Verdi had been compatriots, Verdi and Satie might have been compatriots.
> If Bizet and Verdi had been compatriots, Satie would (still) have been French.
> $\therefore$
> If Bizet, Verdi and Satie had been compatriots, Satie would (still) have been French.

The first premiss is true, because Verdi might have been French; the second, because the hypothesis of Verdi and Bizet's being of the same nationality presumably wouldn't have any bearing on where Satie hails from. However, the conclusion is false, because the possibility that Bizet, Verdie and Satie might all be Italian should not be discarded. From this argument, Ginsberg concludes that Lewis's assumption of transitivity of the relation of being equally close is too restrictive. However, most authors have been reluctant to abandon (ASP), perhaps because the counterexamples to it are admittedly a little *recherché*[3] and the model becomes all too complicated in its absence.

[3]This is a general point. We usually do not use counterfactuals in arguments so as to derive new counterfactuals, but rather to make factual points—most typically in Modus Tollens. In order to assess arguments such as Ginsberg's we therefore often have to shore up our intuitions with theory, with the corresponding threat of circularity. The virtue of Ginsberg's example is that even if we help ourselves to Lewis's system of spheres (which presupposes almost-connectedness), the argument doesn't sound right.

Cotenability from similarity without ASP

$[\![B]\!]$ is cotenable with the counterfactual assumption $[\![A]\!]$ at α iff
$[\![B]\!]_\alpha = [\alpha]_R \vee (\forall \delta \in [\![A]\!]_\alpha)(\exists \beta \in [\![A]\!]_\alpha)(\beta \preceq_\alpha \delta \wedge \forall \gamma(\gamma \preceq_\alpha \beta \supset \gamma \in [\![B]\!]))$

A.3. Kratzer

TC would-counterfactuals Kratzer
$[\![A \mathrel{\Box\!\!\rightarrow} C]\!]^g =$
$\{k \in K \mid (\forall x \in X_{g(k)\cup\{[\![A]\!]\}})([\![A]\!] \in x \supset (\exists y \in X_{g(k)\cup\{[\![A]\!]\}})(x \subseteq y \wedge \bigcap y \subseteq [\![C]\!]))\}$

As long as we can count on the existence of maximally consistent subsets of $g(k) \cup \{[\![A]\!]\}$ containing $[\![A]\!]$[4] (in particular, when $g(k) \cup \{[\![A]\!]\}$ is finite), the above truth condition boils down to $[\![C]\!]$ following from each of them.

TC might-counterfactuals Kratzer
$[\![A \mathrel{\Diamond\!\!\rightarrow} C]\!]^g =$
$\{k \in K \mid (\exists x \in X_{g(k)\cup\{[\![A]\!]\}})([\![A]\!] \in x \wedge (\forall y \in X_{g(k)\cup\{[\![A]\!]\}})(x \subseteq y \supset \bigcap(y \cup \{[\![C]\!]\}) \neq \emptyset)\}$

[4]This requirement is shown in (Lewis, 1981), p.228, to be equivalent to the limit assumption on Stalnaker-Lewis frames.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix B

Resumen en español

En mi disertación doctoral me ocupo del problema de las oraciones condicionales, tal y como ha quedado circunscrito en la tradición analítica de la filosofía actual. Este problema abarca cuestiones tales como el significado de las oraciones condicionales y sus condiciones veritativas (semántica), las peculiaridades de su uso en la práctica comunicativa (pragmática), la justificación canónica de las proferencias de estas oraciones (epistemología), y su comportamiento inferencial (lógica). Una de las tesis principales de esta disertación es que a pesar de su estrecha relación, estas cuestiones deben distinguirse nítidamente, ya que cada una de ellas exige una aproximación metodológica particular.

Introducción

El llamado problema de los condicionales surge con el advenimiento de la lógica clásica moderna y la *barra condicional* con la que Frege codificó los contenidos condicionales en la matemática. En el sistema de la *Conceptografía*, afirmar un contenido condicional consistía simplemente en negar la posibilidad de la afirmación del antecedente y negación del consecuente. Esta sencilla concepción de la condicionalidad fue implementada por Russell en su definición de la *implicación material*, que en el sistema de los *Principia Mathematica* hacía las veces de la barra condicional fregeana.

Si bien las ideas de Frege y Russell cosecharon un enorme éxito en el ámbito de la fundamentación de la matemática, al parecer los fundadores de la lógica matemática no pretendían extrapolar sus resultados al análisis del lenguaje natural. La tendencia a identificar el significado de conectores naturales con el de sus contrapartidas formales es posterior a Frege y Russell y se debe al éxito en la aplicación de la lógica clásica al estudio de algunos fenómenos lingüísticos.

Ahora bien, este último proyecto ha puesto en evidencia las deficiencias del condicional material a la hora de codificar el significado de las oraciones condicionales cotidianas. En primer lugar, los hablantes no tienden a considerar la falsedad del antecedente ni la verdad del consecuente suficientes para la aceptabilidad un condicional—lo cual aparentemente sería el caso si los condicionales de nuestro lenguaje fuesen materiales. Esta intuición es aún más clara en el caso de los condicionales subjuntivos, también llamados contrafácticos, cuyos usos normales presuponen la falsedad de su antecedente. Otra razón para sospechar de la interpretación material de los condicionales naturales es que algunos patrones deductivos válidos para fórmulas con condicionales materiales tienen instancias contraintuitivas si los condicionales materiales se sustituyen por el 'si-entonces' del lenguaje natural.

Una de las estrategias para hacer frente a estas paradojas consiste en desterrarlas al ámbito de la pragmática y mantener la identificación del 'si-entonces' y el condicional material a nivel de significado lingüístico. Es la vía que ha explorado P. Grice en varios de sus trabajos.[1] Según Grice, la intuición de que la verdad del consecuente o la falsedad del antecedente no garantizan la verdad del condicional se puede deber al hecho de que en tales circunstancias, la afirmación de un condicional material sería inapropiada. El llamado *Principio de Cooperación* nos insta a maximizar la informatividad de nuestras contribuciones conversacionales, y dado que el condicional material es estrictamente más débil desde el punto de vista lógico que su consecuente o la negación de su antecedente, afirmar aquél en conocimiento de alguno de éstos sería censurable.

Una estrategia similar, defendida por F. Jackson, identifica las condiciones veritativas del condicional material y natural, pero no su significado lingüístico.[2] La parte adicional del significado del condicional indicativo natural (la llamada *implicatura convencional* a la que da lugar su uso) garantiza la aplicabilidad de la regla inferencial del Modus Ponens—el condicional natural debe ser robusto con respecto a la verdad de su antecedente.

El objetivo tácito de los teóricos de la identidad es defender la tesis de que la lógica clásica es la lógica del lenguaje natural, y que los contraejemplos tienen que ver más con los objetivos comunicativos para los que los condicionales se usan que con su significado lingüístico traducible en términos de condiciones veritativas. Es este último que sostiene las inferencias semánticas.

[1] Véase (Grice, 1989).
[2] (Jackson, 1987)

A pesar de toda su sofisticación, los teóricos de la identidad no consiguen dar cuenta de todas las anomalías del comportamiento del condicional material. Poco o nada tienen que decir sobre la relación de los condicionales indicativos, supuestamente materiales, y los subjuntivos, que se resisten a tal interpretación. Y por si fuera poco, algunos de los argumentos que estos teóricos esgrimen se levantan sobre cimientos sospechosos. Por todo ello, la mayoría de los autores que trabajan en el campo ha concluido que la Tesis de la Identidad es insostenible y que un tratamiento correcto de los condicionales debe romper más radicalmente con la tradición.

Parte I. Las teorías.

En los capítulos 2-5 se exponen y someten al escrutinio las teorías de los condicionales alternativas a la Tesis de la Identidad que más impacto han tenido a lo largo de la segunda mitad del s.XX. Estas teorías incluyen la teoría probabilística de los condicionales indicativos, la teoría basada en las premisas co-sostenibles (cotenability theory) y la teoría basada en la similitud de mundos posibles de los subjuntivos, y las teorías que combinan diferentes elementos de las anteriores, defendidas por C. Gauker, W. Lycan, A. Kratzer y F. Veltman. Un lugar importante lo ocupan también las teorías causales, desarrolladas principalmente para resolver algunos problemas resistentes del análisis de los subjuntivos. La conclusión general de esta parte de la disertación es que la proliferación de los estudios sobre los condicionales no ha redundado en un avance sustancial de nuestra comprensión de los fenómenos relacionados con ellos.

El capítulo segundo está dedicado a la teoría del condicional probabilístico, cuyos orígenes se retrotraen al pensamiento de F. P. Ramsey.[3] Ha sido elaborada y defendida principalmente por E. W. Adams y, más recientemente, D. Edgington, pero ha encontrado amplio eco entre los más diversos autores.[4] La idea principal que sustenta este enfoque es que la probabilidad del condicional indicativo $A \rightarrow C$ equivale a la probabilidad condicional de C dado A (la llamada Tesis de Adams). Este supuesto, al parecer inocente, tiene varias consecuencias sorprendentes. Como ha probado D. Lewis, la Tesis de Adams implica que el contenido de los condicionales indicativos no puede representarse, como es habitual con las proposiciones ordinarias, como un subconjunto del espacio muestral.[5] Esto también requiere modificaciones en la definición de validez deductiva: Adams ha

[3](Ramsey, 1929)

[4]Véase principalmente (Adams, 1975) y (Adams, 1998), (Edgington, 2007) y (Bennett, 2003).

[5]Véase (Lewis, 1976) y (Lewis, 1986b).

sustituido el concepto tradicional por el de la validez probabilística, que difiere de aquél justo cuando entre las premisas se encuentran condicionales indicativos. La interpretación probabilística del condicional indicativo invalida los patrones inferenciales sospechosos respaldados por la Tesis de la Identidad.

Las muchas virtudes de la teoría de Adams, sin embargo, han llevado algunos a sobrestimar su alcance. En vista del éxito de la aplicación de la semántica probabilística a los condicionales indicativos, muchos han concluido que éstos no expresan contenidos proposicionales sobre los que quepa un debate objetivo, sino que su función estriba meramente en transmitir información sobre el sistema de creencias del hablante.[6] Ahora bien, tal doctrina contradice muchos aspectos del uso declarativo de los indicativos y no explica la presencia de oraciones condicionales subordinadas en oraciones principales con distinta fuerza ilocutiva. Además, invita la conclusión de que no hay ninguna relación interesante entre los condicionales indicativos y subjuntivos.[7] La probabilidad por lo tanto no puede ser la última palabra a la hora de explicar los condicionales.

Los capítulos tercero y cuarto están dedicados a teorías de los condicionales subjuntivos. La primera, más antigua, se propone la tarea de delimitar el conjunto de premisas verdaderas y co-sostenibles con el antecedente de un subjuntivo, que junto con él impliquen el consecuente si y sólo si el subjuntivo es verdadero. En otras palabras, estos teóricos pretenden reducir la dependencia contrafáctica a implicación lógica. La virtud de este planteamiento consiste en otorgar a los contrafácticos un papel transparente en las prácticas argumentativas humanas, dispersando la apariencia de irracionalidad o arbitrariedad que a veces suscitan. Sin embargo, el proyecto de especificación de las premisas colaterales se ha topado con obstáculos formidables, sobre todo en lo tocante a la naturaleza de las generalizaciones nomológicas y la direccionalidad de las relaciones causales entre fenómenos. Este último problema, expuesto nítidamente por N. Goodman y conocido por el de nombre de éste, se ha convertido en la piedra de toque de las teorías de subjuntivos, independientemente del tipo de aproximación por el que se decanten.[8]

La teoría actualmente más popular de subjuntivos es sin lugar a dudas la de Stalnaker-Lewis, basada en la similitud entre mundos posibles.[9] Esta teoría nos va a ocupar en el capítulo cuarto. En su versión más sencilla, define un contrafáctico verdadero como aquel cuyo consecuente es verda-

[6] Véase por ejemplo (Bennett, 2003), p.126.
[7] (Gibbard, 1981)
[8] (Goodman, 1947)
[9] (Stalnaker, 1968) y (Lewis, 1973a)

dero en el mundo posible más similar al actual en el que el antecedente es verdadero.

Parece que la principal fuente de atractivo de esta teoría reside en la belleza matemática del sistema al que da lugar. Las propiedades formales que uno atribuya a la relación de similitud determinarán el comportamiento lógico de los contrafácticos, por lo que el sistema ofrece un alto grado de flexibilidad y un amplio foro para debates sobre la lógica contrafáctica 'correcta'. Sin embargo, a lo largo del capítulo intento mostrar que estos debates distan mucho de ser concluyentes, lo cual compromete la verificabilidad de la teoría.

Los cimientos teóricos sobre los que se levanta este análisis de subjuntivos son los conceptos de mundo posible y la similitud entre ellos. Ahora bien, a mi modo de ver, ambos conceptos de por sí tan solo desplazan las cuestiones fundamentales relacionadas con nuestro uso de subjuntivos: la razón de ser y los usos prácticos del vocabulario modal, así como el papel que el pensamiento contrafáctico desempeña en nuestras vidas intelectuales. En particular, los análisis de la relación de similitud que se han propuesto son ad-hoc en un grado extremo y ni siquiera así consiguen sortear todas las objeciones de sus críticos. El resultado es, en las palabras del propio Stalnaker, que esta teoría de los contrafácticos no resuelve ninguna cuestión sustancial, sino que tan solo ofrece un marco formal para su planteamiento.[10] En vista de su vacuidad, es difícil de justificar el grado de atención que esta teoría ha recibido en la bibliografía, sobre todo dado que la mayoría de los autores se ha centrado casi exclusivamente en las cuestiones formales.

Desafortunadamente, teorías poco informativas de los condicionales abundan en el campo. En el capítulo quinto se debaten dos teorías de este tipo: la de W. G. Lycan y la de C. Gauker.[11] Su rasgo común es que su análisis hace un uso desmesurado de variables cuyo valor se adapta flexiblemente al resultado que se desea obtener en cada caso. En Gauker, este papel lo desempeñan los contextos conversacionales, mientras que en Lycan todo el trabajo lo realiza el conjunto de posibilidades contempladas por el hablante. En consecuencia, estas teorías, a pesar de su sofisticación formal, no nos enseñan nada muy relevante sobre nuestro tema de estudio.

Una buena parte del capítulo 5 está dedicada a la semántica de premisas de A. Kratzer.[12] La intuición que guía su tratamiento de los condicionales es que estos siempre funcionan como restrictores de cuantificación, sea

[10](Stalnaker, 1984), p.122
[11](Lycan, 2001) y (Gauker, 2005)
[12]Véase (Kratzer, 2012).

implícita o explícita, estándar o modal. Esta intuición la llevó a postular una modalización implícita en los consecuentes de los condicionales indicativos. Además, Kratzer intentó reconstruir la relación de similitud entre mundos posibles con ingredientes análogos a los de la teoría de la co-sostenibilidad para dar cuenta de los subjuntivos. Para hacer frente a los problemas con los que sin éxito habían lidiado los teóricos anteriores, Kratzer introdujo la noción de 'bulto de pensamientos' (*lump of thought*): algunas premisas no se podían añadir al antecedente por separado, sino tan solo en conjunto con un bulto de otras. Sin embargo, Kratzer hizo un uso mucho más amplio de esa noción de lo que justificaban las intuiciones que habían respaldado su introducción. F. Veltman intentó adaptar la teoría otorgando un papel clave a las generalizaciones nomológicas, pero su intento naufragó en el problema de Goodman.

En vista de los frecuentes fracasos a la hora de evitar problemas como los de Goodman o Tichý, numerosos autores han optado por implementar la dirección de la dependencia causal directamente en el engranaje formal de una teoría de subjuntivos. Mientras que sus rivales pretenden obtener este resultado a partir de ingredientes más elementales, los teóricos causales suelen asumir que la relación de causalidad es irreducible y tiene que figurar entre los inputs de una teoría formal de condicionales contrafácticos: esto les permite evitar los mencionados problemas de un solo golpe. Ahora bien, el precio es que el andamiaje formal de las teorías en sí no contribuya a una mejor comprensión del razonamiento contrafáctico. El recurso directo a la causalidad desplaza todo el interés teórico de una teoría de contrafácticos desde la semántica formal hacia la comprensión del papel que las consideraciones causales juegan en nuestro pensamiento.

Parte II. La propuesta.

En los capítulos sexto y séptimo de esta tesis defiendo que la ruptura con la Teoría de la Identidad se quedó a medio camino. Muchos teóricos que rechazaron la interpretación material de los condicionales se lanzaron en busca de una interpretación alternativa, sin cuestionar los principios metodológicos del antiguo paradigma. En particular, entre éstos figuran los siguientes supuestos:

1. El significado se debe explicar en términos de condiciones veritativas.

2. La semántica formal siempre define el significado de la expresión analizada. Con la proferencia de un condicional, el hablante se compromete con que el estado correspondiente a la intepretación formal del condicional es el caso.

3. El significado, tal y como viene enunciado por la semántica formal, determina la lógica de la expresión analizada.

En el capítulo 6 muestro que el primero de estos supuestos suele ir acompañado con la creencia tácita en la existencia de un vocabulario privilegiado para la tarea de enunciar las condiciones veritativas de expresiones complejas; de otra forma no cabría justificar el interés de este proyecto semántico. No obstante, nada parece garantizar la existencia de tal vocabulario: no hay razón para pensar que la contribución de las expresiones más interesantes de nuestro lenguaje deba de ser reducible a la de otras, en principio más rudimentarias. La vacuidad de muchas teorías en boga quizá se deba a la imposibilidad de reproducir el contenido de un condicional sin recurrir a alguna noción artificial, introducida apelando a nuestra comprensión previa de los condicionales.

En contra de la metodología semántica veritativo-condicional, sugiero que la mejor manera de explicar el significado de una expresión es remitir a la función estándar que ésta realiza en un intercambio conversacional. El significado de muchas expresiones no-veritativo-condicionales solo se puede explicar de esta forma, e incluso en el caso de expresiones que contribuyen a las condiciones de verdad de las proferencias en las que intervienen, este tipo de explicación es muchas veces más iluminador.

Aplicado a nuestro objeto de estudio, un análisis análogo sugiere que las oraciones subordinadas condicionales indicativas sirven para limitar el alcance del acto de habla llevado a cabo a través de la proferencia de la oración principal. La validez de ese acto (que puede ser asertivo, directivo, compromisorio, etc.) está sometido al cumplimiento de la condición enunciada por la oración subordinada.

Un análisis semejante ciertamente roza la trivialidad, pero eso no constituye ninguna excepción entre las explicaciones de significado. Además, pese a ello el análisis permite evitar la frecuente conclusión, basada en el supuesto 2 arriba mencionado y el éxito de la semántica probabilística, de que los condicionales indicativos tan solo expresan hechos acerca del sistema de creencias del hablante y por lo tanto no pueden ser considerados verdaderos ni falsos. Si bien la alta probabilidad condicional de C dado A es esencial para una afirmación veraz de $A \rightarrow C$ y la alta probabilidad subjetiva de A es esencial para una afirmación veraz de A, A y $A \rightarrow C$ no se reducen a afirmaciones sobre las correspondientes probabilidades.

La semántica probabilística, por lo tanto, es una buena herramienta para explorar la dinámica de nuestras creencias condicionales, pero no agota el significado de los condicionales indicativos. Y por lo tanto, tampoco determina la lógica del condicional indicativo. En la vida real, el comporta-

miento inferencial de una expresión no depende tan sólo de su significado, sino que viene parcialmente determinado por el contexto en el que las inferencias se realizan. Normalmente, al hacer una aseveración condicional, tenemos que tener en cuenta el impacto de la eventualidad introducida por el antecedente sobre todas nuestras creencias: es nuestra mejor manera de garantizar que nuestra afirmación tendrá las de resultar verdadera. Ahora bien, en contextos especiales podemos legítimamente decidir proteger una serie de creencias de toda duda, para ver hasta dónde podemos llegar sin tener que desvincularnos de ellas. En esas circunstancias, típicas de un sistema axiomático tradicional, un condicional indicativo será interderivable con la correspondiente disyunción, y su comportamiento por tanto será el de un condicional material. El significado del condicional indicativo por sí solo no puede fijar el marco inferencial en el que va a ser utilizado.

El último capítulo está dedicado a condicionales subjuntivos y tiene un carácter más especulativo. Mi propuesta está basada en las intuiciones que motivan el tratamiento de los subjuntivos como predicciones pasadas, intentando evitar los escollos que acosan tal tratamiento echando mano de los aciertos de las teorías causales.[13] La idea principal es que los seres humanos estructuran su concepción del mundo en escenarios o guiones inferenciales, que tienen un importante componente causal. La inmensa mayoría de lo que llamamos 'hechos' está mediada teóricamente: el concepto de mundo es un concepto en un alto grado construido. Los contrafácticos sobre el pasado, por lo tanto, nos informan acerca de la estructura de los hechos y las expectativas que una persona razonable podía haber tenido con respecto a ellos. Ello a su vez ayuda a que nos forjemos expectativas razonables sobre situaciones análogas en el futuro. Lejos de ser arbitrarios o irrelevantes, los contrafácticos juegan un papel fundamental a la hora de compartir la información sobre lo que cabe esperar en determinadas circunstancias.

No todos los contrafácticos son causales, ni tampoco todos están basados en leyes. De acuerdo con mi propuesta, los condicionales subjuntivos se utilizan para realizar afirmaciones simples sobre relaciones entre proposiciones dentro de un marco inferencial más amplio. Por eso pueden usarse tanto para transmitir contenidos de la teoría de conjuntos como a la hora de explorar las consecuencias hipotéticas de haber tomado una decisión equivocada en algún asunto mundano. El punto en común es que los dos casos conforman escenarios inferencialmente articulados.

[13](Adams, 1975), (Dudman, 1989) y (Edgington, 2007).

Bibliography

Ernest Adams. The logic of conditionals. *Inquiry*, 8:166–197, 1965.

Ernest Adams. Subjunctive and indicative conditionals. *Foundations of Language*, 6:89–94, 1970.

Ernest Adams. *The Logic of Conditionals*. Reidel, Dordrecht, 1975.

Ernest Adams. *A Primer of Probability Logic*. CSLI Publications, Stanford, California, 1998.

J.H. Allen and J.B. Greenough. *New Latin Grammar*. Ginn and Company, London, 1903.

Luis Alonso-Ovalle. Counterfactuals, correlatives and disjunction. *Linguistics and Philosophy*, 32:207–244, 2009.

Ana Arregui. Layering modalities: the case of backtracking counterfactuals. Unpublished manuscript, available at http://aix1.uottawa.ca/ aarregui/ana_arregui_files/layering-modalities.pdf [Last Access: 28/01/15], 2005a.

Ana Arregui. *On the Accessibility of Possible Worlds: The Role of Tense and aspect*. PhD thesis, University of Massachusetts, Amherst, 2005b.

John L. Austin. Cans and ifs. In *Philosophical Papers*, pages 205–232. Oxford University Press, Oxford, 2nd edition, 1970.

Kent Bach. The myth of conventional implicature. *Linguistics and Philosophy*, 22:327–366, 1999.

Jonathan Bennett. Counterfactuals and possible worlds. *Canadian Journal of Philosophy*, IV(2):381–402, 1973.

Jonathan Bennett. Counterfactuals and temporal direction. *Philosophical Review*, 93:57–91, 1984.

Jonathan Bennett. Farewell to the phlogiston theory of conditionals. *Mind*, 97(388):509–527, 1988.

Jonathan Bennett. *A Philosophical Guide to Conditionals*. Oxford University Press, New York, 2003.

Patrick Blackburn, Maarten de Rijke, and Yde Venema. *Modal Logic*. Cambridge Tracts in Theoretical Computer Science. Cambridge University Press, Cambridge, 2004.

Diane Blakemore. *Semantic Constraints on Relevance*. Blackwell, Oxford, 1987.

Diane Blakemore. *Relevance and Linguistic Meaning. The Semantics and Pragmatics of Discourse Markers*, volume 99 of *Cambridge Studies in Linguistics*. Cambridge University Press, Cambridge, 2002.

Robert Brandom. *Making it Explicit*. Harvard University Press, Cambridge, Massachusetts, 1994.

Robert Brandom. *Articulating Reasons*. Harvard University Press, Cambridge, 2001.

Rachael Briggs. Interventionist counterfactuals. *Philosophical Studies*, 160 (1):139–166, 2012.

David Butcher. An incompatible pair of subjunctive conditional modal axioms. *Philosophical Studies*, 44:71–110, 1983.

Roderick Chisholm. The contrary-to-fact conditional. *Mind*, (55):289–307, 1946.

Roderick Chisholm. Law statements and counterfactual inference. *Analysis*, 5(15):97–105, 1955.

Keith DeRose and Richard E. Grandy. Conditional assertions and 'biscuit' conditionals. *Noûs*, 33(3):405–420, 1999.

John Divers. *Possible Worlds*. Routledge, New York, 2002.

P. B. Downing. Subjunctive conditionals, time order, and causation. *Proceedings of the Aristotelian Society*, 59:125–140, 1959.

Oswald Ducrot et al. *Les Mots du Discours*. Minuit, Paris, 1980.

Victor H. Dudman. Conditional interpretation of *If*-sentences. *The Australian Journal of Linguistics*, 4:143–204, 1984.

Victor H. Dudman. Indicative and subjunctive. *Analysis*, 48(3):113–122, 1988.

Victor H. Dudman. Vive la revolution! *Mind*, 98(392):591–603, 1989.

Michael Dummett. Bringing about the past. *Philosophical Review*, 73:338–359, 1964.

Dorothy Edgington. Do conditionals have truth conditions? *Crítica: Revista Hispanoamericana de Filosofía*, 18(52):3–39, 1986.

Dorothy Edgington. On conditionals. In D. M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 14, pages 127–221. Springer, Netherlands, 2nd edition, 2007.

Paul Egré and Mikaël Cozic. If-clauses and probability operators. *Topoi*, 30:17–29, 2011.

Herbert B. Enderton. *Elements of Set Theory*. Academic Press, London, 1977.

John Etchemendy. *The Concept of Logical Consequence*. Harvard University Press, Harvard, 1990.

Niall Ferguson. Virtual history: Towards a 'chaotic' theory of the past. In Niall Ferguson, editor, *Virtual History: Alternatives and Counterfactuals*, pages 1–90. Basic Books, 1999.

Kit Fine. Critical notice on (Lewis, 1973a). *Mind*, 84(1):451–8, 1975.

Kit Fine. Counterfactuals without possible worlds. *Journal of Philosophy*, 109(3):221–246, 2012.

Michael Franke. The pragmatics of biscuit conditionals. In Maria Aloni, Paul Dekker, and Floris Roelofsen, editors, *Proceedings of the 16th Amsterdam Colloquium*, pages 91–96. University of Amsterdam, 2007.

Gottlob Frege. Compound thoughts. *Mind*, 72(285):1–17, 1963.

Gottlob Frege. Begriffsschrift, a formula language, modeled upon that of arithmetic, for pure thought. In Jean van Heijenoort, editor, *From Frege to Gödel: A Source Book in Mathematical Logic*, pages 1–82. Harvard University Press, 1967.

María J. Frápolli. *The Nature of Truth. An Updated Approach to the Meaning of Truth Ascriptions*, volume 29 of *Logic, Epistemology and the Unity of Science*. Springer, 2013.

María J. Frápolli and Neftalí Villanueva. Minimal expressivism. *dialectica*, 66(4):471–487, 2012.

David Galles and Judea Pearl. An axiomatic characterization of causal counterfactuals. *Foundations of Science*, 3(1):151–182, 1998.

Cristopher Gauker. What is a context of utterance. *Philosophical Studies*, 91: 149–172, 1998.

Cristopher Gauker. *Conditionals in Context*. MIT Press, Cambridge, 2005.

Allan Gibbard. Two recent theories of conditionals. In W.L. Harper, R. Stalnaker, and G. Pearce, editors, *Ifs*, pages 211–247. D. Reidel, Dordrecht, 1981.

Matthew L. Ginsberg. Counterfactuals. *Artificial Intelligence*, 30(1):35–79, 1986.

Nelson Goodman. The problem of counterfactual conditionals. *Journal of Philosophy*, 5(44):113–128, 1947.

Nelson Goodman. Parry on counterfactuals. *Journal of Philosophy*, 54:442–5, 1957.

Nelson Goodman. *Fact, Fiction and Forecast*. Harvard University Press, Cambridge, Massachusetts, 4th edition, 1983.

Herbert Paul Grice. The causal theory of perception. Reprinted in his (1989), pp.224-247, 1961.

Herbert Paul Grice. Indicative conditionals. Reprinted in his (1989), pp.58-87, 1967.

Herbert Paul Grice. Logic and conversation. Reprinted in his (1989), pp.22-40, 1975.

Herbert Paul Grice. Further notes on logic and conversation. Reprinted in his (1989), pp.41-57, 1978.

Herbert Paul Grice. *Studies in the Way of Words*. Harvard University Press, Cambridge, 1989.

William H. Hanson. Indicative conditionals are truth-functional. *Mind*, 397:53–72, 1991.

W.L. Harper, R. Stalnaker, and G. Pearce, editors. *Ifs*. Reidel, Dordrecht, 1981.

Irene Heim and Angelika Kratzer. *Semantics in Generative Grammar*. Blackwell, Oxford, 1998.

Hans G. Herzberger. Counterfactuals and consistency. *Journal of Philosophy*, 76(2):83–88, 1979.

Alan Hájek. Probabilities of conditionals - revisited. *Journal of Philosophical Logic*, 18:423–428, 1989.

Alan Hájek. The fall of 'Adams' thesis'? *Journal of Language, Logic and Information*, 21:145–161, 2012.

Alan Hájek and Ned Hall. The hypothesis of the conditional construal of conditional probability. In Ellery Eells and Brian Skyrms, editors, *Probability and Conditionals. Belief Revision and Rational Decision*, pages 75–111. Cambridge University Press, Cambridge, 1994.

Ted Honderich. Causes and *if p, even if x, still q*. *Philosophy*, 57:291–317, 1982.

Sabine Iatridou. The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 30(2):231–270, 2000.

Frank Jackson. A causal theory of counterfactuals. *Australasian Journal of Philosophy*, 55(1):3–21, 1977.

Frank Jackson. *Conditionals*. Blackwell, Oxford, 1987.

Tony Judt. *Postwar*. Penguin Press, New York, 2005.

Makoto Kanazawa, Stefan Kaufmann, and Stanley Peters. On the lumping semantics of counterfactuals. *Journal of Semantics*, 22:129–151, 2005.

Stefan Kaufmann. Causal premise semantics. *Cognitive Science*, 37:1136–1170, 2013.

William Kneale and Martha Kneale. *The Development of Logic*. Clarendon Press, Oxford, 1962.

Tapio Korte, Ari Maunu, and Aho Tuomo. Modal logic from Kant to possible worlds semantics. In Leila Haaparanta, editor, *The Development of Modern Logic*, pages 516–550. Oxford University Press, New York, 2009.

Angelika Kratzer. What 'must' and 'can' must and can mean. *Linguistics and Philosophy*, 1:337–355, 1977. Reprinted in her (2012), pages 1-20.

Angelika Kratzer. Conditional necessity and possibility. In R. Bäuerle, U. Egli, and A. von Stechow, editors, *Semantics from Different Points of View*, pages 117–147. Springer, Berlin, Heidelberg and New York, 1979.

Angelika Kratzer. Partition and revision: the semantics of counterfactuals. *Journal of Philosophical Logic*, 10:201–216, 1981a.

Angelika Kratzer. The notional category of modality. 1981b. Reprinted with modifications in (Kratzer, 2012), pp.27-69.

Angelika Kratzer. An investigation of the lumps of thought. *Linguistic and Philosophy*, 12:607–653, 1989.

Angelika Kratzer. Conditionals. 1991. Reprinted with modifications in (Kratzer, 2012), pp.86-108.

Angelika Kratzer. Constraining premise sets for counterfactuals. *Journal of Semantics*, 22:153–158, 2005.

Angelika Kratzer. *Modals and Conditionals*. Oxford University Press, Oxford, 2012.

Richard Ned Lebow. What's so different about a counterfactual? *World Politics*, 52:550–585, 2000.

Clarence I. Lewis. Implication and the algebra of logic. *Mind*, 21(84): 522–531, 1912.

Clarence I. Lewis. *A Survey of Symbolic Logic*. University of California Press, Berkeley, 1918.

David Lewis. *Counterfactuals*. Harvard University Press, Cambridge, Massachusetts, 1973a.

David Lewis. Counterfactuals and comparative possibility. *Journal of Philosophical Logic*, 2(4):418–446, 1973b.

David Lewis. Causation. *The Journal of Philosophy*, 70(11):556–567, 1973c.

David Lewis. Adverbs of quantification. In Edward L. Keenan, editor, *Formal Semantics of Natural Language*, pages 178–188. Cambridge University Press, Cambridge, 1975.

David Lewis. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 86(3):297–315, 1976.

David Lewis. Possible-world semantics for counterfactual logics. *Journal of Philosophical Logic*, 6:359–363, 1977.

David Lewis. Counterfactual dependence and time's arrow. *Noûs*, 13(4): 455–476, 1979.

David Lewis. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10:217–234, 1981.

David Lewis. *On the Plurality of Worlds*. Basil Blackwell, Oxford, 1986a.

David Lewis. Probabilities of conditionals and conditional probabilities II. *Philosophical Review*, 95(4):581–589, 1986b.

Konrad Lorenz. Kants Lehre vom Apriorischen im Lichte gegenwärtiger Biologie. In Irenäus Eible-Eibesfeld, editor, *Das Wirkungsgefüge der Natur und das Schicksal des Menschen*, pages 82–109. Piper, München, 1978.

Edward N. Luttwak. If Bosnians were dolphins. *Commentary*, 96:27–32, October 1993.

William G. Lycan. A syntactically motivated theory of conditionals. In Peter A. French, Theodore E. Uehling, and Howard K. Wettstein, editors, *Causation and Causal Theories: Midwest Studies in Philosophy*, volume 9, pages 437–455. University of Minnesota, 1984.

William G. Lycan. *Real Conditionals*. Oxford University Press, New York, 2001.

John L. Mackie. Counterfactuals and causal laws. In R. J. Butler, editor, *Analytical Philosophy*, pages 66–80. Blackwell, Oxford, 1962.

John L. Mackie. *Truth, Probability and Paradox*. Oxford University Press, Oxford, 1973.

Richard B. Miller. Moderate modal realism. *Philosophia*, 28(1-4):3–38, 2001.

Peter Milne. Bruno de Finetti and the logic of conditional events. *British Journal for the Philosophy of Science*, 48:195–232, 1997.

Donald Nute. Counterfactuals and the similarity of worlds. *Journal of Philosophy*, 72(21):773–778, 1975.

Donald Nute and Charles B. Cross. Conditional logic. In D.M. Gabbay and F. Guenthner, editors, *Handbook of Philosophical Logic*, volume 4, pages 1–98. Kluwer Academic Publishers, 2 edition, 2001.

William Tuthill Parry. Reëxamination of the problem of counterfactual conditionals. *Journal of Philosophy*, 54(4):85–94, 1957.

Judea Pearl. *Causality: Models, Reasoning and Inference*. Cambridge University Press, 2000.

John L. Pollock. *Subjunctive Reasoning*. Reidel, Dordrecht, 1976.

Willard van Orman Quine. *Methods of Logic*. Holt, Rinehart and Winston, New York, 2nd edition, 1966.

Frank P. Ramsey. Universals of law and fact. Published posthumously in his (1978), pages 128-132, 1928.

Frank P. Ramsey. General propositions and causality. Published posthumously in his (1978), pages 133-151, 1929.

Frank P. Ramsey. *Foundations*. Routledge, London, 1978.

François Recanati. *Literal Meaning*. Cambridge University Press, 2004.

Nicholas Rescher. Belief-contravening suppositions. *Philosophical Review*, (70):176–196, 1961.

Nicholas Rescher. *Conditionals*. MIT Press, London, 2007.

Steven Rieber. Conventional implicatures as tacit performatives. *Linguistics and Philosophy*, 20(1):50–72, 1997.

Adam Rieger. Conditionals are material: the positive arguments. *Synthese*, 190:3161–3174, 2013.

Bertrand Russell. *Principles of Mathematics*. Routledge Classics, Oxon, 2010.

Bertrand Russell and Alfred North Whitehead. *Principia Mathematica*. Cambridge University Press, Cambridge, 2nd edition, 1964. Reprinted to ∗56.

Gilbert Ryle. *The Concept of Mind*. Hutchinson, London, 1949.

David H. Sanford. *If P then Q: Conditionals and the Foundations of Reasoning*. Routledge, London, 1989.

Katrin Schulz. 'If you'd wiggled A, then B would've changed': Causality and counterfactual conditionals. *Synthese*, 179:239–251, 2011.

Katrin Schulz. Fake tense. Available at http://www.uva.nl/over-de-uva/organisatie/medewerkers/content/s/c/k.schulz/k.schulz.html [Last Access: 19 November 2014], 2012.

Brian Skyrms. Adams conditionals. In Ellery Eells and Brian Skyrms, editors, *Probability and Conditionals. Belief Revision and Rational Decision*, pages 13–26. Cambridge University Press, Cambridge, 1994.

Robert Stalnaker. A theory of conditionals. In N. Rescher, editor, *Studies in Logical Theory*, number 2, pages 98–112. Basil Blackwell, 1968.

Robert Stalnaker. Probability and conditionals. *Philosophy of Science*, 37(1): 64–80, 1970.

Robert Stalnaker. Indicative conditionals. *Philosophia*, 5(3):269–286, 1975.

Robert Stalnaker. Letter to van Fraassen. In W.L. Harper and C.A. Hooker, editors, *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, volume I, pages 302–306. D.Reidel, Dordrecht, 1976.

Robert Stalnaker. A defense of conditional excluded middle. In William L. Harper, Robert Stalnaker, and Glenn Pearce, editors, *Ifs*, pages 87–104. Reidel, Dordrecht, 1981.

Robert Stalnaker. *Inquiry*. MIT Press, Cambridge, Massachusetts, 1984.

Peter Frederick Strawson. 'If' and '⊃'. In *Philosophical Grounds of Rationality: Intentions, Categories, Ends*.

Patrick Suppes. Some questions about Adams' conditionals. In Ellery Eells and Brian Skyrms, editors, *Probability and Conditionals. Belief Revision and Rational Decision*, pages 5–11. Cambridge University Press, Cambridge, 1994.

Philip E. Tetlock and Aaron Belkin. Counterfactual thought experiments in world politics. In Philip E. Tetlock and Aaron Belkin, editors, *Counterfactual Thought Experiments in World Politics: Logical, methodological and psychological perspectives*, pages 3–38. Princeton University Press, 1996.

Pavel Tichý. A counterexample to the Stalnaker-Lewis analysis of counterfactuals. *Philosophical Studies*, 29:271–273, 1976.

Pavel Tichý. A new theory of subjunctive conditionals. *Synthese*, 37:433–457, 1978.

Pavel Tichý. Subjunctive conditionals: Two parameters vs. three. *Philosophical Studies*, 45:147–179, 1984.

Bas van Fraassen. Probabilities of conditionals. In W.L. Harper and C.A. Hooker, editors, *Foundations of Probability Theory, Statistical Inference and Statistical Theories of Science*, volume I, pages 261–301. D.Reidel, Dordrecht, 1976.

Frank Veltman. Prejudices, presuppositions and the theory of counterfactuals. In J. Groenendijk and M. Stokhof, editors, *Amsterdam Papers in Formal Grammar*, volume 1, pages 248–281. University of Amsterdam, Amsterdam, 1976.

Frank Veltman. *Logics for Conditionals*. PhD thesis, University of Amsterdam, Amsterdam, 1985.

Frank Veltman. Making counterfactual assumptions. *Journal of Semantics*, 22:159–180, 2005.

Frank Veltman. Counterfactuals, the standard theory. ILLC, University of Amsterdam: lecture notes, 2012.

Ken Warmbrōd. A defense of the limit assumption. *Philosophical Studies*, 42:53–66, 1982.

Michael Williams. *Problems of Knowledge*. Oxford University Press, Oxford, 2001.

Timothy Williamson. *Knowledge and its Limits*. Oxford University Press, Oxford, 2000.

Deirdre Wilson and Dan Sperber. Relevance theory. In Laurence R. Horn and Gregory Ward, editors, *Handbook of Pragmatics*, pages 607–632. Blackwell Publishing, Oxford, 2004.

Michael Woods. *Conditionals*. Oxford University Press, Clarendon Press, 1997.