ASPECTS OF THE MEANING AND USE OF CONDITIONALS

A DISSERTATION
SUBMITTED TO THE DEPARTMENT OF LINGUISTICS
AND THE COMMITTEE ON GRADUATE STUDIES
OF STANFORD UNIVERSITY
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Stefan Kaufmann
August 2001

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Stanley Peters
(Principal Advisor)

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

David Beaver

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Cleo Condoravdi

I certify that I have read this dissertation and that in my opinion it is fully adequate, in scope and quality, as a dissertation for the degree of Doctor of Philosophy.

Peter Sells

Approved for the University Committee on Graduate Studies:

Abstract

Conditional sentences pose a problem for semantic accounts in terms of truth conditions. On the one hand, evidence about the validity of inference patterns suggests that there is a close relationship between conditionals and conditional probability. On the other hand, while probability is typically interpreted as the probability that a proposition is true, that connection has proven elusive in the case of conditionals. Some probabilistic accounts accommodate these obstacles by concluding either that conditionals do not have truth conditions, or that the usual connection between truth and probability does not hold for them. Both of these approaches remain problematic; moreover, definitions of the probability of conditionals directly in terms of conditional probability do not extend easily to complex and embedded conditionals, a limitation of coverage imposed solely by the probabilistic calculus.

In this dissertation I develop a way to reinstate the connection between conditional probability and probability of truth by adopting a new definition of the latter. Building on previous work on the formal aspects of the problem, I suggest that conditionals denote random variables which, unlike the denotations of other sentences, may take values between 1 ("True") and 0 ("False"); the probability of a conditional is the expectation of its values. I explore the rationale for such values and the predictions made by the theory in detail for three common types of conditionals: epistemic, predictive, and counterfactual. Under the proposed theory, conditionals are interpreted in models of the evolution of objective chance and subjective probabilities, both of which are the sources of the values of different kinds of conditionals. The time of evaluation thus plays a crucial part in determining those values.

The theory predicts values for complex and embedded conditionals as well, but the expectation of these values in a given situation does not always accord with intuitions about what the probability of the sentence should be. I show that this problem can be successfully addressed by making the interpretation sensitive to causal dependencies. This addition opens up new insights into the relationship between counterfactual conditionals and their epistemic and predictive counterparts: They are equivalent but not equiprobable.

Acknowledgements

I am grateful to my advisers and many others who shared their insights and influenced my thinking in innumerable ways. Stanley Peters has been a patient teacher, benevolent boss and unyielding critic. He encouraged me to pursue my chosen topic when I was still groping for ways to even state it, and our many consultations were instrumental in shaping this work. His insistence on clarity and rigor and his ability to quickly derail a shaky argument with a few incisive questions pushed me along and helped set my head straight many a time. David Beaver, Cleo Condoravdi and Peter Sells all had extremely helpful comments on both contents and form of my drafts at the many stages they went through. Among the others who helped me at various points with conversations and comments on my ideas were Ernest Adams, Johan van Benthem, Martin van den Berg, Dick Crouch, Mary Dalrymple, Paul Dekker, Hideyuki Nakashima, Robert van Rooy, Henriette de Swart, Robert Stalnaker, Elizabeth Traugott, Frank Veltman and Ed Zalta.

I was fortunate to be surrounded by fine colleagues and fellow students; to name but a few at the expense of others: Ray Flournoy, John Fry and Dominic Widdows at the Center for the Study of Language and Information; Michael Foster, Bob Rama, Bob Tierney, Chris Scott and Ethan Segal in the Department of Asian Languages; Ash Asudeh, Brady Clark, Martina Faller, Dave McKercher, Hanjung Lee, Sarah Roberts and Shiao-Wei Tham in the Department of Linguistics. My special, heartfelt thanks go to Cathryn Donohue for her enduring friendship, and to my wonderful office mates, Devyani Sharma and Ida Toivonen. Not to forget Hugh, Cal, Roge, Ray, Ben, Gene and the others who made paragliding in the Monterey Bay Area such a pleasurable pastime.

My deepest gratitude is due to my family for their support, tolerance and love—to my parents, Brigitte and Heinz, to Michael, and to Misa for her unwavering faith and commitment.

Contents

Abstract **iv**

Acknowledgements **vi**

1 Introduction **1**

1.1 Preliminaries 2

1.2 Classifying conditionals 4

1.2.1 Indicative *vs.* counterfactual 4

1.2.2 Open *vs.* hypothetical 7

1.2.3 Predictive *vs.* non-predictive 11

1.2.4 Conclusion 16

1.3 Some theories of conditionals 17

1.3.1 The material conditional 18

1.3.2 The variably strict conditional 22

1.3.3 Conclusion 27

1.4 The scope of this work 28

1.4.1 Condition, concession, and contrast 28

1.4.2 Indirect conditionals 29

1.4.3 Direct conditionals 30
1.4.4 Syntactic peculiarities 31

2 Probability **34**
2.1 Introduction 35
2.2 Sources of Uncertainty 35
2.2.1 Non-determinism 35
2.2.2 Ignorance 36
2.2.3 Vagueness and ambiguity 37
2.2.4 Methodological remarks 39
2.3 Conditional probability: The Thesis 40
2.3.1 Probabilistic inference 42
2.3.2 The Thesis as a pragmatic condition 43
2.3.3 The agenda 45
2.4 Implementation 47
2.4.1 Probabilities of propositions 47
2.4.2 Probabilities of truth-functional sentences 48
2.5 Probabilities of conditionals 51
2.5.1 The proposition '*if A then C*' 52
2.5.2 Intermediate values 56
2.6 Questions to be addressed 57

3 Time, truth, and knowledge **60**
3.1 Introduction 61
3.2 Time 61

3.2.1 Intuitions 62
3.2.2 Historical necessity 64
3.3 Truth in time 67
3.3.1 Temporal reference 67
3.3.2 Truth and settledness 70
3.3.3 Truth and future facts 71
3.3.4 Chance 72
3.3.5 Metaphysical remarks 73
3.4 Knowledge in time 74
3.4.1 Partiality 74
3.4.2 Knowledge about history 75
3.4.3 Known and settled facts 77
3.4.4 Credence 78
3.4.5 Epistemological remarks 79
3.5 Probabilities of sentences 79
3.6 Summary 81

4 The values of conditionals **82**

4.1 Introduction 83
4.2 Predictive conditionals 83
4.2.1 The consequent 83
4.2.2 The antecedent 85
4.2.3 The conditional 87
4.3 Epistemic conditionals 88

4.3.1 Objective values . . . 88
4.3.2 Subjective values . . . 89
4.3.3 Conditional credence and conditional chance . . . 90
4.3.4 Subjective value assignment . . . 91
4.3.5 Conclusion . . . 93
4.4 Right-nested conditionals . . . 94
4.4.1 The Import-Export Principle . . . 95
4.4.2 Constancy . . . 100
4.4.3 Interpreting compounds of conditionals . . . 101
4.4.4 The values at non-antecedent worlds . . . 110
4.5 Summary . . . 117

5 Counterfactuals and causality **118**
5.1 Introduction . . . 119
5.2 Counterfactuals . . . 119
5.2.1 Past predominance . . . 119
5.2.2 Overall similarity . . . 121
5.2.3 Posterior facts . . . 122
5.2.4 Causal independence . . . 127
5.3 Causality . . . 133
5.3.1 What causal relations relate . . . 134
5.3.2 Particular and general causation . . . 135
5.3.3 Deterministic causation of probabilities . . . 136
5.3.4 Implementation . . . 137

5.4 Applications 140
5.4.1 Right-nested conditionals 140
5.4.2 Conjunctions of conditionals 145
5.4.3 Counterfactuals: Prior propensity 146
5.4.4 Import-Export revisited 151
5.5 Summary 152

6 Conclusion **153**

A Free compounds and embeddings **155**
A.1 Introduction 155
A.2 Informal characterization 156
A.3 Implementation 156

Bibliography **162**

Chapter 1

Introduction

1.1 Preliminaries 2

1.2 Classifying conditionals 4

1.2.1 Indicative *vs.* counterfactual 4

1.2.2 Open *vs.* hypothetical 7

1.2.3 Predictive *vs.* non-predictive 11

1.2.4 Conclusion 16

1.3 Some theories of conditionals 17

1.3.1 The material conditional 18

1.3.2 The variably strict conditional 22

1.3.3 Conclusion 27

1.4 The scope of this work 28

1.4.1 Condition, concession, and contrast 28

1.4.2 Indirect conditionals 29

1.4.3 Direct conditionals 30

1.4.4 Syntactic peculiarities 31

1.1 Preliminaries

Few expressions of the English language rival simple '*if* ... *then* ... ' sentences on the score of attracting prolonged and intense interest and yet stubbornly resisting analysis. Some of the most basic questions about them continue to puzzle and polarize the large community of authors writing on their linguistic and philosophical aspects: In how many different guises do conditionals come? How many different conditionals are behind those guises? Do any or all conditionals have truth values, or in any case "objective values," some or all of the time? How are those objective values related to those of other sentences? How are beliefs in conditionals related to beliefs in other sentences? Is the assertion of a conditional the assertion that it is true?

I will present an analysis which addresses these questions with respect to three "guises" of conditionals, which I label *epistemic*, *predictive*, and *counterfactual*, illustrated by the examples in (1.1a–c).

(1.1) a. If you struck the match, it lit. (epistemic)
b. If you strike the match, it will light. (predictive)
c. If you had struck the match, it would have lit. (counterfactual)

By way of preview, I will give the following answers to the above questions: Predictive and counterfactual conditionals generally have objective values, albeit not necessarily truth values in the usual sense. Epistemic conditionals have truth conditions, which may however fail to produce values in certain cases. There is a common semantic core behind all three types which provides the basis for relating their values to those of other sentences and in addition allows for the identification of the the belief in a conditional with the belief that it is true.

In the bigger picture, the main goal and contribution of this dissertation is to reconcile two theoretical approaches, both of which have their respective merits: probabilistic theories of *inference* and logical theories of *truth*. The former offer an elegant account of the role of conditionals in reasoning. The latter model logical relations between conditionals, their constituents and other sentences, but require special additional stipulations to prevent them from making the wrong predictions about inference. A long-standing problem in the relationship between the two is that while probabilistic accounts of simple sentences can in general be built upon truth-theoretical interpretations, there is no obvious way to extend this connection to conditionals. Building on previous work on the technical aspects of this problem, I show that there is a somewhat less obvious, but nonetheless highly illuminating way to make the connection.

The discussion of the probabilistic approach, its motivation and the problems involved in relating the probabilities of conditionals to their truth values starts in Chapter 2. There I will also outline the basics of the truth conditions I adopt for conditionals to solve the problem and make the connection to probability. It will not become clear at that point, however, that this solution has any significance besides the fact that it allows for the desired connection to be made. In particular, the definitions call for the assignment of rather unusual "truth values" which in turn depend on the probability. Chapter 2 will stop short of demonstrating that those values are anything more than dummies introduced for mathematical convenience.

In Chapter 3, I will address this last objection by offering an interpretation of the underlying probability as objective *chance*. This affords an account of the otherwise puzzling fact that the truth conditions make reference to the probability. Objective chances are determined at any given time by the history up to that time. The interpretation of conditionals, too, is temporally located, and their values vary with time. Thus reference to the probability in the truth conditions is in fact no more unusual than reference to the time of evaluation. In Chapter 3 I develop the formal apparatus required to model this dependence both of chance and the values of conditionals on time, as well as to relate objective chance to subjective *credence*, a speaker's estimate of the former and the source of the subjective probabilistic values of sentences.

Chapter 4 spells out the truth conditions in more detail for a large class of simple and right-nested conditionals. It demonstrates the role of time and belief for simple conditionals as well as the values assigned by the approach about larger compounds and embeddings. It has been pointed out before that those values are intuitively wrong. I will argue that judgments as to why they are wrong are guided by intuitions about counterfactuals and causal relations.

These two notions are discussed in their own right in Chapter 5. Empirical facts about counterfactuals clearly demonstrate the need for a role of causality in the interpretation. I will discuss the nature of causality and its implementation in the model, and finally its application to the interpretation to both conditional and non-conditional sentences.

The remainder of this introduction serves some preliminary discussions: In Section 1.2, I will make my classification of conditionals more precise and relate it to commonly used alternatives. Section 1.3 gives a brief overview of some prominent semantic accounts of conditionals, their motivations and shortcomings. Finally, in Section 1.4 I mention some types and uses of conditionals which are beyond the scope of this work.

1.2 Classifying conditionals

In the last section I gave three example sentences (1.1a–c), each illustrating one of the classes of conditionals I distinguish. Here I will elaborate on my choice of these three categories by discussing some distinctions that have been found useful before, their morphological correlates as well as intuitions about their semantic properties.

1.2.1 Indicative *vs.* counterfactual

The labels "indicative" and "counterfactual" are used by most philosophers and many linguists to draw a major boundary between classes of conditionals. The term "subjunctive" is sometimes used instead of "counterfactual" and generally meant to be coextensive with the latter (but not always—cf. the quotation from Lewis (1973) on page 7). With an eye towards the discussion that follows, I use examples from Quirk et al. (1985, p. 1091) for illustration.[1]

(1.2) If Colin is in London, he is undoubtedly staying at the Hilton.

(1.3) a. They would be here with us if they had the time.

b. If you had listened to me, you wouldn't have made so many mistakes.

Those authors who comment on their own use of these terms often hasten to add that they are not supposed to be taken literally. Indeed, both "subjunctive" and "counterfactual" are misnomers:

As a morphological term, "subjunctive" is useless in making the distinction. To the extent that English has preserved the fading distinction between indicative and subjunctive, the use of the latter is optional in conditionals, restricted to formal style and, most importantly, possible in sentences of both types (Quirk et al., 1985). Thus (1.4) is grouped with (1.2) despite the fact that it contains the subjunctive form *'be'*.[2]

(1.4) If any person be found guilty, he shall have the right to appeal.

On the other hand, both of (1.5a,b) are grouped with (1.3), even though the verb forms are subjunctive and indicative, respectively.[3]

(1.5) a. If I were rich, I would buy you anything you wanted.

b. If I was rich, I would buy you anything you wanted.

For more discussion on the problems with the label "subjunctive," see Dudman (1988) and Bennett (1988). Dudman speculates that the persistence of its use in

[1]Quirk et al. do not use the terms "indicative" or "counterfactuals," and their distinction does not draw the same line (see the discussion below).

[2]Examples (1.4) and (1.5) are from (Quirk et al., 1985, pp. 1012 and 158, respectively).

[3]Some grammarians disagree with this formally motivated terminology. Edgington (1995) notes that Fowler (1950) would treat the past indicative used in (1.5b) as ambiguous between indicative and subjunctive.

the literature is due to an "inchoate but reassuring Latin Prose Theory" according to which "just those conditionals are indicative (subjunctive) which translate into indicative (subjunctive) conditionals in Latin" (Dudman, 1988, p. 119).

Nor is it advisable to take the term "counterfactual" literally. It makes reference to an implicature commonly observed in sentences like (1.3) to the effect that the antecedent is presumed false by the speaker. Indeed, this generalization is often borne out: (1.3b) suggests that "they" do not have time, and (1.3c), that the listener did not listen to the speaker.

On the other other hand, it was noticed early on by Anderson (1951) that sentences of the same type can be used to argue *for* the truth of the antecedent: (1.6) is most naturally used to support the claim that the patient did in fact use arsenic, as that would be the best explanation for the observed symptoms.

(1.6) If he had taken arsenic, he would have shown just these symptoms.
[those which he in fact shows]

Anderson's observation has been mentioned frequently in the literature on conditionals (Karttunen and Peters, 1979; Barwise, 1986; Comrie, 1986), but the term "counterfactual" has survived in spite of this problem.

Given the inadequacy of the terms, their use is often accompanied by disclaimers depriving them of their content. Chisholm (1946) uses "subjunctive" and "contrary-to-fact" interchangeably, although "neither term is adequate." (p. 482) Lewis (1973) concedes:

> My title 'Counterfactuals' is too narrow for my subject. I agree, but I know no better." (p. 3)

Somewhat more uncertain is the status of sentences like (1.7).[4]

(1.7) If he changed his opinion, he would be a much more likeable person.

[4]Example (1.7), too, is from Quirk et al. (1985).

Sentences like this are generally not set apart from "counterfactuals," which seems plausible given that (1.7) suggests that "he" is not going to change his opinion, hence will remain disliked. However, to Lewis (1973) it is subjunctive, but *not* counterfactual:

> The title 'Subjunctive Conditionals' would not have delineated my subject properly. ... [T]here are subjunctive conditionals pertaining to the future, like *'If our ground troops entered Laos next year, there would be trouble'* that appear to have the truth conditions of indicative conditionals, rather than of the counterfactual conditionals I shall be considering. (p. 4)

Thus in Lewis' terms, "subjunctive" and "counterfactual" are not coextensive, unlike in the use of many other authors.

1.2.2 Open *vs.* hypothetical

In the linguistic literature, the classification of conditionals is terminologically more diverse and no less controversial than in philosophy. Quirk et al. (1985) take as the most important criterion the attitude of the speaker to the truth of the antecedent and accordingly define the following classes:[5]

open: noncommittal with respect to the fulfillment of the condition, hence also that of the matrix clause;

also known as *real, factual, neutral*

hypothetical: conveying the speaker's belief that the condition will not be, is not, or was not fulfilled, thus implying the "probable or certain falsity of the proposition expressed by the matrix clause."

also known as *closed, unreal, rejected, nonfactual, counterfactual, marked*

[5]Quirk et al. note that the literature knows a number of different labels for the same classes, but do not associate specific authors or writings with those (Note [a], page 1092).

The classification is illustrated by (1.8) for open and (1.9) for hypothetical conditionals.

(1.8) If Colin is in London, he is undoubtedly staying at the Hilton.

(1.9) a. If he changed his opinion, he would be much more likeable.
b. They would be here with us if they had the time.
c. If you had listened to me, you wouldn't have made so many mistakes.

These classes coincide with the indicative (1.8) and subjunctive (1.9) conditionals of Section 1.2.1, hence with indicative and counterfactual conditionals for those authors for whom "subjunctive" and "counterfactual" are coextensive. The "hypotheticality" in the sentences (1.9a–c) consists in the following "implications" (Quirk et al.'s term) conveyed by the wording of the protases:

(1.10) a. He very probably won't change his opinions.
b. They presumably don't have the time.
c. You certainly didn't listen to me.

In general, hypothetical conditionals convey the speaker's belief that the antecedent

(1.11) a. *will* not be fulfilled (contrary to expectation)
b. *is* not fulfilled (contrary to assumption)
c. *was* not fulfilled (contrary to fact)

The parenthesized formulations on the right were used by Dahl (1997) and Dancygier (1998).

1.2.2.1 Backshift

The division just outlined has the advantage of combining the epistemic criterion of the speaker's attitude towards the truth of the antecedent with the morphological phenomenon of *backshift*: The clauses embedded under *'if'* in sentences like (1.9a–c) refer to events whose temporal location relative to the speech time is not reflected in their tense in the usual way. Quirk et al. (1985) summarize this anomaly in Table 1.1.

Backshift is generally acknowledged to be an important feature in the classification. As the examples in (1.9) illustrate, it consists in the addition of a "layer" of past morphology to the verb phrase in the protasis, signalling the "contrariness" (to expectation, assumption, or fact) of the antecedent. Although one may take this non-temporal use of morphemes normally associated with tense to be no more than a coincidence (cf. Portner (1992)), various proposals have been made to explain the two in terms of a semantic commonality. Such a semantic explanation would be desirable in light of the fact that past morphology and counterfactuality are intimately related in a great variety of unrelated languages (James, 1982; Fleischman, 1989; Iatridou, 2000, and references therein).

Table 1.1: Backshift (Quirk et al., 1985)

	conditional clause	matrix clause
present and future reference	Past *'If I were younger,'*	Past Modal *'I would study Classical Creek.'*
past reference	Past Perfective *'If I had seen you,'*	Past Perfective Modal *'I would have invited you home.'*

One such suggestion, traced back to Joos (1964) by Dahl (1997), is to treat both pastness and counterfactuality as specifications of a general meaning of "distance" or "remoteness." Some proponents of this approach (Iatridou, 2000) take this to be the underspecified meaning of past morphology, of which both past and counterfactuality are instantiations in context. Others (Fleischman, 1989; James, 1982) treat the temporal meaning as basic and "remoteness" as that element of it which supports its *metaphorical* extension to the non-temporal domain.

Yet another school of thought (Dudman, 1984a; Dahl, 1997) seeks to give the additional past morpheme its usual temporal meaning and *derive* the counterfactual interpretation. The basic idea is that the protases of these conditionals do not merely refer to (non-actual) *facts*, but alternative *courses of events* ("fantasies" in Dudman's terms), which depart from the actual history at a time prior to that of the hypothesized state of affairs. According to these accounts, the fact that this point of departure (called the "cross-over point" by Dudman and the "choice point" by Dahl) lies in the past from the perspective of the hypothetical state of affairs necessitates the additional past morpheme.

This point of departure from history is an implicit parameter of the interpretation; its precise location is usually not expressed, hence it is a potential source of vagueness. In any case, it is a time at which it was still possible that the antecedent would come true, and hypothetical conditionals are evaluated by *extrapolating* from that time under the assumption that it *is* (or was, or will be) true. According to Tedeschi (1981),

> [w]e evaluate counterfactual conditional sentences as if we returned to the past and looked at possible futures with respect to that past.

Thus according this view, counterfactuals are *predictions*. This intuition is shared by many other authors (Downing, 1959; Adams, 1975; Ellis, 1978; Lewis, 1979; Thomason and Gupta, 1981; Dudman, 1984a, 1994; Strawson, 1986; Bennett, 1988; Mellor, 1993; Dancygier, 1998, and others).

In addition to the intuitive appeal of an account along these lines, Dahl (1997) notes that it also explains the invariable use of past forms of *'will'* in the apodoses of hypothetical conditionals. For instance, (1.12a) asserts that at (and before) the point of departure, i.e. the point at which it became unlikely that he would change his opinion, the sentence *'He will be a more likeable person'* was true under the supposition that he would change his opinion; similarly for (1.12b,c).

(1.12) a. If he changed his opinion, he'd be a more likeable person.

If he changes his opinion, he will be a more likeable person.

b. They would be here with us if they had time.

They will be here with us if they have time.

c. If you had listened to me, you wouldn't have made so many mistakes.

If you listen to me, you won't make so many mistakes.

Indeed, there is a clear similarity between (1.13a–c) and their "open" counterparts: Intuitively, the hypothetical prediction is true if and only if the open conditional *was* true at or just before the point of departure. This calls into question the status of the line between "open" and "hypothetical" conditionals as separating two semantically quite different kinds of conditionals.

1.2.3 Predictive *vs.* non-predictive

In the last section I pointed out a semantic parallelism between hypothetical conditionals and their open counterparts. To illustrate, the sentences in (1.13) are familiar from Adams (1970).

(1.13) a. If Oswald does not kill Kennedy, somebody else will.

b. If Oswald had not killed Kennedy, somebody else would have.

c. If Oswald did not kill Kennedy, somebody else did.

In Quirk et al.'s classification, (1.13a,c) are open conditionals while (1.13b) is hypothetical. In alternative terms, (1.13a,c) are indicative while (1.13b) is counterfactual. The user of (1.13b) conveys that Oswald did in fact kill Kennedy, whereas the user of (1.13a,c) takes no stance on that question.

The presence or absence of an implicature concerning the truth of the antecedent is the decisive criterion in the classifications considered so far, hence the drawing of a major division division between (1.13a,c) on one side and (1.13b), on the other.

1.2.3.1 The stand-alone test

It is clear that there is a semantic difference between (1.13b) and (1.13c): One can consistently deny the former while accepting the latter. What is more controversial is the question as to whether this semantic difference runs deeper than that between (1.13a) and (1.13c), as the division into open and hypothetical (or indicative and counterfactual) sentences would suggest. Dudman (1984a) pointed out that the tense marking in the protases of sentences like (1.13a) is anomalous in ways similar to the backshift of (1.13b), although it does not involve the addition of a tense morpheme. Taken in isolation, (1.14a) is an odd thing to say.

(1.14) a. ??Oswald does not shoot Kennedy.
b. Oswald won't / is not going to shoot Kennedy.

(1.14b) is much more natural in isolation, but on the other hand, embedding (1.14b) in the protasis results in (1.15), which is not synonymous with (1.13a).

(1.15) If Oswald won't / isn't going to shoot Kennedy, somebody else will.

There are two ways of interpreting (1.15). Under one reading, what is at issue is not whether Oswald does or does not end up shooting Kennedy, but rather whether he currently *intends* to do so. The second reading, discussed, for instance, by Palmer (1983) and Comrie (1985), involves a causal link from the consequent to the antecedent; (Palmer, 1983, p. 242) suggests that conditionals of this sort are elliptical and can be "spelled out" as in (1.16a). The intended reading is brought out even more clearly by the paraphrase in (1.16b).

(1.16) a. If, if somebody else shoots Kennedy Oswald won't, somebody else will.
b. If someone else's shooting Kennedy will prevent Oswald from shooting him, someone else will shoot him.

Neither of these two readings is that of (1.13a). The fact remains that a clause concerning the question as to whether Oswald will or won't shoot Kennedy appears without the morpheme *'will'* when embedded under *'if'*. Various proposals have been made to account for this phenomenon. Comrie (1985) treats it as the result of a syntactic transformation, assuming that *'will'* is covertly present but deleted overtly. Dudman (1984b) takes a somewhat different approach, assuming that *'will'* is never present in the first place. Instead, the protasis is not a clause in its own right, but an adverbial modifier of the modal in the apodosis. Dancygier (1998) treats it as just another instance of backshift, effected by canceling out, as it were, the future reference and an added past. A third possibility is to take (1.15a) literally, hence the protasis of (1.13a) as expressing the supposition that (1.15a) is true.[6]

Whatever the explanation of this temporal anomaly is, and regardless of whether it is morphologically transparent, it is absent in sentences like (1.13c). The clause in (1.17) has the same temporal reference in isolation that it has in the the protasis of (1.13c).

(1.17) Oswald did not shoot Kennedy.

Dudman argues in a series of papers (1984a; 1984b; 1989; 1991; 1994) that this difference reflects the most important distinction within the class of open (indicative) conditionals:[7]

> [(1.18a)] and [(1.18b)] are as different grammatically as it is possible for two sentences to be, whereas [(1.18b)] and [(1.18c)] are grammatically congruent. (1989, p. 591)

[6]I said above that such "naked futures" are an odd thing to use in isolation. However, they are not ungrammatical, albeit used only in certain contexts. As (Comrie, 1985, p. 118) notes, this use is felicitous when the sentence is "scheduled" to become true. For instance, a screenwriter can felicitously use (1.15a) to explain the further development of the plot. (Edgington, 1997, p. 98) credits the "Almighty" with the authority to make statements like (1.15a) about his or her intentions, but dismisses that option as irrelevant in the case of conditional protases. It is not obvious to me that it is irrelevant, and I will speculate some more on the question in Chapter 4.

[7]Dudman, as well as Funk (below), uses different examples.

(1.18) a. If Oswald does not shoot Kennedy, somebody else will.

b. If Oswald did not shoot Kennedy, somebody else will.

c. If Oswald did not shoot Kennedy, somebody else did.

Furthermore, according to Dudman, sentences like (1.18a) are semantically similar to hypothetical conditionals in important respects; thus the temporal reference of the protasis marks the major fault-line in the landscape of conditionals.

1.2.3.2 The nature of uncertainty

The same proposal was made independently in the linguistic literature. Funk (1985) relates the difference in temporal reference to one in the status of the *facts* referred to. All of (1.18a–c) are used under uncertainty about the truth of the antecedent: Oswald may or may not shoot (1.18a) or have shot (1.18b,c) Kennedy. However, the uncertainty has different sources:

> In the case of [(1.18a)] the uncertainty is largely due to the fact that the state-of-affairs described and predicated does not yet exist, i.e., is still subject to manifestation (so that it cannot be affirmed or denied—it is unverifiable) at the moment of the sentence being uttered. In [(1.18b,c)], however, the state-of-affairs *does* exist at the time of speaking (either in the positive or negative sense—it is 'manifested' and could thus be verified), but the speaker has not got enough information (or is otherwise not disposed) to be sure about it and hence to affirm or deny it. Accordingly, the meaning of the conditioning frame can be said to vary from "if it happens that ... " to "if it is true that ... " (pp. 375–376)

Funk adduces evidence that the distinction is real, drawing on data from Latin and Greek which I will not discuss here. It is reassuring, however, to know that there is such cross-linguistic evidence, since in English the distinction lacks clear formal reflexes. While it would not be natural to use the protasis of (1.18a) in isolation, the same does not hold for (1.19a).

(1.19) a. If John is in his office, we'll have the meeting there.

b. If John is in his office tomorrow at 11:00, we'll have the meeting there.

c. If John is in his office already, he came early this morning.

The condition in (1.19a) may concern either whether John is in his office "now" or whether we will find him there at some relevant later time. This is a general property of *stative* predicates as in (1.19a), and it is for this reason that they, unlike the *eventive* ones as in (1.18), do not display the difference Dudman uses as his criterion.

Instead, the interpretation is guided by contextual factors such as the salience of a specific time the sentence is about, or by temporal indexicals as in (1.19b,c). In the presence of such disambiguating expressions, stative predicates behave like eventive ones: (1.20) is again unnatural.[8]

(1.20) ??John is in his office tomorrow at 11:00.

1.2.3.3 Terminology

This new differentiation is accompanied by further terminological complications. Funk (1985) uses the the terms "open" and "closed" differently from Quirk et al. (1985): His "open" ones include only a proper subset of theirs and in addition all those that Quirk et al. call "hypothetical." Dancygier (1998) makes the same distinction as Funk, but uses the term "predictive" (Funk's "open") and "non-predictive" (Funk's "closed.") Dudman prefers to label as "conditionals" only Funk's "open" ones, calling "closed" ones "hypothetical"—not to be confused with Quirk et al.'s "hypotheticals." Table 1.2 summarizes these conflicting uses, applied to the sentences in (1.21).

(1.21) a. If Oswald did not shoot Kennedy, somebody else did.

b. If Oswald did not shoot Kennedy, somebody else will.

c. If Oswald does not shoot Kennedy, somebody else did.

[8] In addition, (1.20a) is felicitous in the same restricted circumstances as eventive predicates (cf. Footnote 6 on page 13).

d. If Oswald does not shoot Kennedy, somebody else will.

e. If Oswald were not to shoot Kennedy, somebody else would.

f. If Oswald did not shoot Kennedy, somebody else would.

g. If Oswald had not shot Kennedy, somebody else would have.

Table 1.2: Terminologies in the classification of conditionals

<table>
<tr><th></th><th>(1.22a,b)</th><th>(1.22c,d)</th><th>(1.22e)</th><th>(1.22f,g)</th></tr>
<tr><td rowspan="2">Lewis[9]</td><td colspan="2">indicative</td><td colspan="2">subjunctive</td></tr>
<tr><td colspan="3"></td><td>counterfactual</td></tr>
<tr><td>Quirk et al.</td><td colspan="2">open;
real, factual, neutral</td><td colspan="2">hypothetical;
closed, unreal, rejected, nonfactual, counterfactual, marked</td></tr>
<tr><td rowspan="2">Bennett[10]</td><td>straight cond.</td><td colspan="3">corner cond.</td></tr>
<tr><td colspan="2">straight cond.</td><td colspan="2">corner cond.</td></tr>
<tr><td>Dudman</td><td>hypothetical</td><td colspan="3">undeclarative; conditional</td></tr>
<tr><td>Funk</td><td>closed</td><td colspan="3">open</td></tr>
<tr><td>Dancygier</td><td>non-predictive</td><td colspan="3">predictive</td></tr>
<tr><td>Kaufmann</td><td>epistemic</td><td>predictive</td><td colspan="2">counterfactual</td></tr>
</table>

1.2.4 Conclusion

My own choice, given in Table 1.2, is motivated by the arguments for both of the distinctions discussed in the preceding sections. I take both the distinction between indicative and counterfactual conditionals and that between predictive and

[9] The two rows indicate Lewis' grammatical (upper) and semantic (lower) distinctions.

[10] The labels 'straight' and 'corner' used by Bennett (1988, 1995) are inspired by the shape of the connectives '$\rightarrow$' and '$>$', respectively. Bennett was anxious not to use terms which carry too much connotational baggage. He changed his mind about the classification between 1988 and 1995.

non-predictive ones to be useful and intuitively real. The long-standing and sometimes heated philosophical debate on the issue of where "the" line should be drawn only underscores the point that there are arguments for both.

In the account I will develop, the indicative-counterfactual distinction is treated as one largely of tense, although it does not follow that present counterfactuals are equivalent to earlier indicatives. The difference between epistemic and predictive conditionals lies in presence or absence, respectively, of what I call a *presumption of settledness*. Chapter 3 discusses this notion in more detail.

1.3 Some theories of conditionals

No less controversial than the taxonomy of conditionals itself is the question of its semantic correlate.

The distinction between indicative and counterfactual conditionals is often associated with fundamentally different semantic treatments. Adams (1970) introduced the examples in (1.13), repeated here as (1.22a,b):

(1.22) a. If Oswald did not Kennedy, then someone else did.

b. If Oswald had not killed Kennedy, then someone else would have.

From the observation that "[(1.22a)] is probably true while [(1.22b)] may very well be false," Lewis (1973) concludes:

> [T]here really are two different sorts of conditional; not a single conditional that can appear as indicative or as counterfactual depending on the speaker's opinion about the truth of the antecedent. (p. 3)

Not everyone agrees. Edgington (1995) aptly suggests that the difference may be "more like the difference between mature cheddar and freshly-made cheddar than the difference between chalk and cheese" (p. 239). Strawson (1986) uses a minimal pair like (1.23a,b) to make the same point:[11]

(1.23) a. Remark made on November 21, 1963:
"If Oswald does not kill Kennedy, somebody else will."

b. Remark made on November 23, 1963:
"If Oswald had not killed Kennedy, somebody else would have."

Strawson writes:

> It seems obvious that about the least attractive thing one could say about the *difference* between these two remarks is that it shows that, or even that it is partly accounted for by the fact that, the expression *'if ... then ...* ' has a different meaning in one remark from the meaning which it has in the other. (p. 230)

Despite these intuitions, the idea that the semantic distinction is a profound one has proven persistent in the literature. Edgington (1995) conjectures that this may be due to the influence of the Fregean material conditional on the one hand and the importance of counterfactual reasoning in the theory of science (e.g. in the analysis of dispositions,) on the other.

1.3.1 The material conditional

The standard connectives of propositional logic—conjunction, disjunction, negation, and material conditional—are defined as functions taking truth values as arguments and returning truth values. Most speakers agree that the standard definitions of the first three pose no challenges to linguistic intuition, at least as far as pure truth-functional meaning, not inferences arising from implicatures are concerned.

[11]Strawson uses different sentences.

A few basic assumptions are sufficient to show that the *'if... then ... '*-construction *has to* be translated as the material conditional if it is a truth-functional connective like the others. One such argument was given by Dorothy Edgington:

Fact 1 (Edgington (1986))
Let $\mathcal{F} : \{0,1\}^2 \mapsto \{0,1\}$ be the truth function corresponding to 'if φ then ψ', whose arguments are the truth values $V(\varphi)$ and $V(\psi)$ ("0" for "false" and "1" for "true.") Assume in addition that (i) sentences of the form 'if φ and ψ then φ' are tautologous, and (ii) conditionals can be false. Then $\mathcal{F}$ is the truth function of the material conditional.[12]

However, the material conditional is not an adequate representation of natural-language conditionals. In the next section I list some well-known facts to show this.

1.3.1.1 Inference patterns

Each of the following four subsections will start with a valid inference pattern involving the material conditional, then present a linguistic example to show that the same pattern is not valid with the "natural" conditional.

Strengthening the antecedent.

(1.24) $$\frac{A \supset C}{AB \supset C}$$

[12]*Proof.* By Assumption (i), the sentence *'if p and q then p'* is a tautology, i.e., $\mathcal{F}(V(\text{'}p \text{ and } q\text{'}), V(p)) \equiv 1$. There are four cases:

	$V(p)$	$V(q)$	$V(\text{'}p \text{ and } q\text{'})$	$\mathcal{F}(V(\text{'}p \text{ and } q\text{'}), V(p)) = 1 = \ldots$
(a.)	1	1	1	$\mathcal{F}(1,1)$.
(b.)	1	0	0	$\mathcal{F}(0,1)$.
(c.)	0	1	0	$\mathcal{F}(0,0)$.
(d.)	0	0	0	$\mathcal{F}(0,0)$.

Cases (a.–d.) exhaust three of the four possible combinations of arguments of $\mathcal{F}$. By Assumption (ii), conditionals can be false, hence $\mathcal{F}(1,0) = 0$. □

Another argument to the same effect is due to Gibbard (1981). As I will discuss at length in Chapter 4, I do not accept one of Gibbard's premises, *viz.* that *'if p then 'if q then r''* is equivalent to *'if p and q then r'*.

However: Far from contradicting (1.25a), (1.25b) is a perfectly natural continuation. Thus the two sentences in (1.25c) are consistent.[13]

(1.25) a. "If I install a better alarm system," Griliches said, "that is an improvement in the quality of my life, and therefore a decline in inflation.

b. But if the burglars learn how to trick this alarm system, that is a rise in price, because the quality advantage will be eroded. nyt961217.0474

c. $A \rightarrow C$ and $AB \rightarrow \overline{C}$.

Contraposition.

(1.26) $$\frac{A \supset C}{\overline{C} \supset \overline{A}}$$

However: (1.27a) may be true while (1.27b) is false. One may assent to the former while rejecting the latter.

(1.27) a. If you're a high achiever, it takes a long time to get recognition." nyt961130.0128

b. ??If it takes a short time to get recognition, you're a low achiever.

c. $A \rightarrow C$. But *not* $\overline{C} \rightarrow \overline{A}$.

Vacuous truth.

(1.28) $$\frac{\overline{A}}{A \supset C}$$

However: The truth of (1.29a) does not ensure the truth of (1.29b).

(1.29) a. The flood crest won't reach the levels projected.

b. ??If the flood crest reaches the level projected, much of the city will be under water.

c. $\overline{A}$. But *not* $A \rightarrow C$.

[13]Examples with labels of the form 'nyt—.—' are attested in the New York Times Corpus. The numbers refer to their location in the corpus.

Hypothetical Syllogism.

(1.30)
$$\frac{B \supset C \quad A \supset B}{A \supset C}$$

However: It is possible to assent to both of (1.31a,b), yet reject (1.31c).

(1.31) a. If I quit my job, I can't afford my apartment.

b. If I win a Million, I'll quit my job.

c. ??If I win a Million, I can't afford my apartment.

d. $B \to C$ and $A \to B$. But *not* $A \to C$

It has been pointed out (Adams, 1975; Kratzer, 1986, and elsewhere) that counterexamples to Transitivity typically rely on changes in the conversational background and involve premises which would not in fact be tenable simultaneously with respect to the *same* background. In (1.31), too, this can be revealed by changing the order of the premises: (1.32a,b) together are much less plausible than (1.31a,b), a fact which is attributable to the tendency to interpret (1.32b) as the *modally subordinated* (1.32c).

(1.32) a. If I win a Million, I'll quit my job.

b. ??If I quit my job, I can't afford my apartment.

c. ??If I *win a Million and* quit my job, I can't afford my apartment.

This is in line with the more general fact that the inference illustrated in (1.33) does indeed seem valid. This, too, needs to be explained.

(1.33)
$$\frac{A \to B \quad AB \to C}{A \to C}$$

1.3.1.2 Conclusion

The inference patterns listed above show that conditionals are not suitably represented by the material conditional. Together with the fact that the latter is the only truth function that could plausibly represent them, it follows that conditionals are not truth-functional connectives.

The invalid inference patterns all illustrate one problem with the material conditional: It is true "too easily"—it follows from premises from which the "natural" conditional does not follow. On the positive side, however, it seems to be right in its predictions of *falsehood*. An early and generally undisputed statement of this observation is found in Ramsey (1929):

> *'If p then q'* can in no sense be true unless the material implication $p \supset q$ is true; but it generally means that $p \supset q$ is not only true but deducible or discoverable in some particular way not explicitly stated. (p. 156)

In other words, when p is true and q is false, *'if p then q'* is false; otherwise, *'if p then q' may be* true. Attempts to provide a more adequate analysis ought to account for the former and elaborate on the latter.

1.3.2 The variably strict conditional

An alternative approach treats conditionals as quantifiers over possible worlds. The *strict conditional* $\Box(p \supset q)$ is a modalized version of the material conditional: In standard possible-worlds models, it is true if and only if the material conditional is true at every world. The *variably* strict conditional is characterized by a restriction of the scope of the modal operator to a subset of all possible worlds.

The variably strict conditional has been employed both to formalize a description of the cognitive process involved in evaluating conditionals that is known as the "Ramsey Test," and in analyzing the logic of counterfactuals. While the two implementations share their intuitive motivation, they are best kept apart because originally at least, the former was concerned with *beliefs* or conditions on *use* while the latter was aimed at giving *truth* conditions.

1.3.2.1 The Ramsey Test

Ramsey (1929) wrote what is without doubt the most influential footnote in the whole body of literature on conditionals:

> 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, $\overline{q}$' 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 $\overline{p}$ for certain, the question ceases to mean anything to him except as a question about what follows from certain laws or hypotheses. (p. 155)

Stalnaker (1968) generalizes this rule to encompass counterfactuals as well. If the antecedent is known to be *false*, it is incompatible with the stock of knowledge and adding it will lead to inconsistency. Hence, Stalnaker suggests that in those cases some further consistency-restoring adjustments must be made. Stalnaker's reformulation reads as follows:

> First, add the antecedent (hypothetically) to your stock of beliefs; second, make whatever adjustments are required to maintain consistency (without modifying the hypothetical belief in the antecedent); finally, consider whether or not the consequent is now true. (p. 44)

1.3.2.2 Indicative conditionals

The Ramsey Test inspired a variety of formal accounts besides Stalnaker's own (Kratzer, 1979, 1981; Veltman, 1985; Gärdenfors, 1988). Here I will focus on the treatment given by these accounts to *indicative* conditionals. For those, the question of which "adjustments" are required to maintain consistency does not arise, since it is assumed that indicative conditionals are only used when the antecedent is compatible with the stock of beliefs.

The connection to universal quantification over possible worlds is straightforward: Let Γ be a set of sentences representing a "stock of beliefs." Semantically, Γ corresponds to the set K_Γ of worlds at which all the sentences in Γ are true. Adding an antecedent A to Γ eliminates from K_Γ those worlds at which A is false. A conditional $A \to C$ is supported by Γ if and only if C is true in *all* the worlds in the resulting set $K_{\Gamma\cup\{A\}}$.[14]

Note that since the background against which it is interpreted is a "stock of beliefs" (or, alternatively, the information state in which a hypothetical agent finds himself,) the test decides whether the sentence is *believed* (or *supported* by the available information,) not whether it is *true*.

The lack of a definition of truth in the usual sense may itself be taken as a serious shortcoming. But even without it, it is easy to see that the invalid inference patterns remain problematic, for the belief in a conditional *'if A then C'* is supported by a stock of beliefs Γ if and only if the material conditional $A \supset C$ is. In particular, if A is false at all worlds in K_Γ, the conditional is vacuously true. Furthermore, the conditional is false whenever K_Γ contains but one world at which is A is true and C is false.

However, this failure to account for the inference patterns need not be held against the general idea that the interpretation of conditionals involves *some* quantifier. It is the *universal force* of the quantification that reintroduces the problem. A different quantifier may prove more suitable; in Chapter 2 I discuss a probabilistic alternative. There I will also return to the second issue raised in this section, concerning the connection between *support* by a stock of beliefs and *truth*.

[14]Veltman (1985) does not represent stocks of beliefs as sets of worlds; however, he, too, interprets conditionals in terms of universal quantification.

1.3.2.3 Counterfactuals

Quantification over possible worlds has been used to analyze the logic of counterfactuals as well. The major proposals to this effect are due to Stalnaker (1968) and Lewis (1973). Although the two theories differ in detail, they are similar in relevant respects.[15]

Counterfactuals are typically (though not always) used when their antecedent is false. As a consequence, the truth values of either antecedent or consequent are of little relevance to their truth. (1.34) may be true or false (or indeterminate,) but the fact that I did not drop the vase and that it did not break does not decide that question.

(1.34) If I had dropped the vase, it would be broken.

Instead, (1.34) is evaluated by examining alternative worlds where I did drop the vase. If it broke in all of those worlds, the counterfactual is true; otherwise it is false.

The invalidity of the inference illustrated above with indicative conditionals carries over to counterfactuals; this is easy to check by substituting the corresponding forms. This problem is addressed by restricting the set of antecedent-worlds to those that are most *similar* to the actual one: (1.34) is true despite the fact that there are possible worlds where the floor is covered with layers of blankets, for those worlds are less similar to ours than those at which, as in ours, it is made of concrete. Thus it is possible for (1.34) to be true while (1.35) is false.[16]

(1.35) If I had dropped the vase and the floor had been covered with layers of blankets, the vase would be broken.

[15] The differences and commonalities are discussed in many of places, including Lewis (1973); Stalnaker (1981); Gärdenfors (1988).

[16] Restricting the accessibility of worlds in this way corresponds to adding to the antecedent a set of true sentences that are consistent with it. In this form, the idea goes back to Goodman (1947) and is formalized in *premise semantics* (Kratzer (1979, 1981); also Veltman (1985)). For a comparison, see Lewis (1981b).

Stalnaker's and Lewis' theories come apart in the underlying assumptions about similarity, hence the truth values of counterfactuals. Lewis maintains that for a world w and antecedent A, there may be any number (none, one, finitely or infinitely many) of "nearest" antecedent-worlds that are tied in their similarity to w. Stalnaker maintains that there is at most one.[17] Both agree that each world is most similar to itself, with the consequence that at a world at which the antecedent is true, the truth of a conditional comes down to the truth of the material conditional.

While universal quantification is central to the *truth* of counterfactuals in Lewis' theory, in Stalnaker's it enters the picture only at the level of epistemic *support*, in a similar way as that discussed above for indicative conditionals. The underlying intuition is again the Ramsey Test: A set K of worlds[18] is transformed into one in which the antecedent A holds throughout by selecting for each world w in K the A-world that is most similar to w. The selection is restricted to A-worlds in K whenever there are any. If there are, then the result is the subset of A-worlds in K, as before: Each A-world in K selects itself, and each non-A-world selects some A-world.[19]

If there are no A-worlds in K, the selection of alternatives leads to worlds outside K. This, Stalnaker (1975) argues, is the central semantic correlate of the difference between indicative and counterfactual conditionals.[20]

The Stalnaker-Lewis approach is successful in the analysis of the *logic* of counterfactuals. The use of the notion of similarity between worlds in regulating accessibility correctly invalidates the problematic inference patterns and provides a host of additional insights.

[17]Formally, *exactly* one: Stalnaker reserves an artificial "impossible" world for cases in which there is no accessible antecedent-world.

[18]Stalnaker (1975) calls this is the "context set" (representing the presupposed background information,) rather than a "stock of beliefs."

[19]Thus in that case, which Stalnaker associates with the interpretation of indicative conditionals, the monotonicity properties of the operation of hypothesizing are left intact and Strengthening of the Antecedent is still incorrectly predicted valid.

[20]He writes: "I take it that the subjunctive mood in English and some other languages is a conventional device for indicating that presuppositions are being suspended, which means in the case of *conditional* statements, that the selection function is one that may reach outside of the context set." (p. 70)

It is, however, not fully satisfactory as a semantic theory. In stipulating the unanalyzed similarity relation between worlds as a part of the model, it provides "a model theory with room for wired-in answers to various counterfactual questions" (Jeffrey, 1991, p. 162). But although Lewis (1973) maintains that facts about which antecedent-worlds are most similar to a given world *are* facts about that world, too (p. 69), no connection is made to its more "mundane" facts. Van Fraassen (1976) puts it as follows:

> To the question what principles govern deductive reasoning involving conditionals, Stalnaker and Lewis give exact answers. But the validity of an argument does not depend on whether its premises are true; and indeed, Stalnaker and Lewis have not notably increased our ability to decide whether particular conditionals are true or false. (p. 266)

1.3.3 Conclusion

The failure of the material conditional to correctly predict the validity of inference patterns involving conditionals rules out the possibility of an analysis of conditionals as truth-functional connectives. The alternative in terms of the variably strict conditional is intuitively more appealing, but faces two open questions:

1. What is the *force* of quantification?
2. How is the *domain* of the quantification determined?

The former is not satisfactorily addressed in terms of universal quantification, which in the theories of indicative conditionals I discussed fails to account for the problematic inference patterns. The latter concerns theories of counterfactuals: Can the set of most "similar" worlds to the world of evaluation be somehow determined with reference to the facts of that world alone?

In Chapter 2 I will give arguments for a probabilistic alternative to the universal quantification of the variably strict conditional. This solves the problem of the inference patterns, but raises the probabilistic analog of the second question above. Chapter 3 addresses that question.

1.4 The scope of this work

In the terminology of Quirk et al. (1985), the data considered here are *direct* conditionals of the form *'if protasis, (then) apodosis'*, where both protasis and apodosis are fully-fledged clauses. To clarify, I will list in this section various kinds of sentences that are *not* included in the class and hence beyond the scope of this work.

1.4.1 Condition, concession, and contrast

Quirk et al. (1985) place conditionals in the larger class of expressions of "relationships of condition, concession, and contrast." There is no straightforward correspondence between the formal properties of sentences and their place in this rough taxonomy.

Semantically, conditionals are described as conveying "that the situation in the matrix clause is contingent on that in the subordinate clause" (p. 1088). *'If-then'*-sentences, often summarily called "conditionals," can be put to uses other than to express this relationship. Thus (1.36) and (1.37) are contrastive and concessive, respectively.

(1.36) If Benchley and his associates lacked earnestness, they did not lack interest in money. nyt940711.0181

(1.37) If much of this previously unseen work seems simply frivolous, every bit of it is entertaining. nyt961210.0477

These sentences are best paraphrased as (1.38) and (1.39). I cannot detect a relationship of conditional "contingency" in them.

(1.38) While/although Benchley and his associates lacked earnestness, they did not lack interest in money.

(1.39) Although much of this previously unseen work seems simply frivolous, every bit of it is entertaining.

Thus sentences like (1.36) and (1.37) are beyond the scope of this work.

1.4.2 Indirect conditionals

Quirk et al. (1985) divide conditionals into *indirect* and *direct* ones and call the former "peripheral." While it fits their overall characterization of the meaning of conditionals as asserting a "contingency" relationship, that relationship asserted by indirect conditionals holds on a pragmatic meta-level, between speech acts rather than sentence denotations. (1.40) does not assert that the *truth* of the speaker's claim depends on whether she "may say so"; rather, her *use* of the claim is made to depend on its appropriateness in the situation. The conditional clause in (1.40a) is a "hedging" device whose function is purely pragmatic.

(1.40) "Most people don't really know me at all, if I may say so," she says, sitting at her desk at the American Enterprise Institute, with its arresting view of the Jefferson Memorial. nyt940816.0322

Similarly, in (1.41) it is the use of a metaphor that is made dependent upon the listener's (in this case, the reader's) readiness to accept its applicability.

(1.41) The idea is clean, fresh romance in a natural, outdoor setting. Calvin Klein meets Eddie Bauer, if you will. nyt940809.0339

Examples like (1.42) are indirect conditionals as well. It, too, is used under uncertainty regarding the applicability of a metaphor.

(1.42) If the Internet is the world's largest library, it lacks a decent card catalogue. nyt961211.0235

Other conditional speech acts occur as well, illustrated in (1.43).

(1.43) If a steeple is a phallic symbol, is there a Freudian impulse? nyt961205.0264

(1.44) If it tastes good, spit it out it's probably bad for you. nyt940719.0248

This dissertation is not concerned with sentences of this kind.

1.4.3 Direct conditionals

As opposed to these "meta-uses," the relationship asserted by the *direct* variety is characterized by Quirk et al. as follows:

> [T]he truth of the proposition in the matrix clause is a consequence of the fulfillment of the condition in the conditional clause. (p. 1088)

This description calls for precisification. For now, however, it is sufficient to get an impression of the kind of examples discussed here. What is at issue is a relation between the *truth* of the antecedent and that of the consequent.

Within this class of direct conditionals, a further distinction must be made between *generic* and *specific* cases. I will not deal with the former, examples of which are (1.45) and (1.46).

(1.45) "Usually if a search and seizure motion is successful at a preliminary hearing there is no trial," Bell said. nyt940701.0009

(1.46) Today, if Pounds wears jeans, she dresses them up with a red tuxedo vest and pearl earrings. nyt941025.0058

Although generic conditionals are intuitively closely related to the sentences I deal with here, I do not attempt here to cover both classes.

The sentences I do deal with in this dissertation, then, are exemplified by the following:

(1.47) "If the RPF comes here and threatens the population, we will open fire against them without any hesitation," the colonel said, "and we have the means." nyt940704.0241

(1.48) "If you wanted to set up a system to be unaccountable, this is what you would do," says San Francisco Assistant District Attorney John Dwyer, whose office is examining charges regarding the bridge district. nyt940706.0016

(1.49) If God had meant Harvard professors to appear in People magazine, She wouldn't have invented The New York Review of Books. nyt940701.0228

(1.50) "If this bill were to have passed, poor quality produce could have reached the marketplace," said Carla Agar, a spokeswoman for the state Department of Food and Agriculture. nyt940709.0215

This somewhat arbitrarily chosen list is not representative of the full range of data. More relevant sentences will be discussed below.

1.4.4 Syntactic peculiarities

The class of expressions of conditionality is formally diverse. In this work I only consider sentences of the form '*if* A *then* C' in which *A* and *C* occur in that order and are full clauses. This excludes certain sentences on purely formal grounds.

1.4.4.1 Elliptical clauses

Sentences in which one of the clauses is incomplete, e.g. elliptical dependent on surrounding material, are beyond the scope of this work. Such sentences include (1.51) through (1.53).

(1.51) They agreed that the state's fixed funds (such as those set aside for the roads) could be plundered for as much as half a billion dollars in cash if needed. nyt940701.0059

(1.52) The United States has proposed a global ban, but would allow some production of material for civilian use, if placed under international safeguards. nyt940701.0274

(1.53) Morales wholeheartedly agreed, saying he would show his books to the Parks Department if asked. nyt940702.0176

1.4.4.2 *'If'*-less conditionals

In certain contexts "conditionality" arises without the aid of the marker *'if'*, as in (1.54) through (1.56):

(1.54) "What is happening to them is something beyond their control," Clark said. "If a juror goes to the supermarket, he will be exposed. He goes to the gym, he will be exposed. He goes to church on Sunday, he will be exposed. There is no way to avoid it." nyt941024.0431

(1.55) You drink—you drive—you're ours (Sticker seen on a police car)

(1.56) Buy one—get one free (Frequently used in product promotions)

Modal subordination (Roberts, 1989), the interpretation of sentences as dependent on an implicitly or contextually given condition, is another environment which gives rise to conditional interpretations. (1.57) is an example of an elaborate construction with "layers" of conditionality, introduced with explicit marking by *'if'* but in force beyond the sentences containing them and "stacked up" recursively.

(1.57) If a girl gets pregnant anyway, you bring her into one of the many new family planning clinics you've established by borrowing against the massive savings that will inevitably result from the dismantling of the welfare state. You explain her abortion or adoption options. You offer her career counseling, education assistance, and if she insists on having the baby anyway, you mobilize the vastly beefed-up family-law unit in the district attorney's office and go after this deadbeat dad who fathered a child and casually walked away. If the father has a job, you garnish his wages. If he drives a car you take away his license until he agrees to pay child support. If the father neither has a job nor a car, you place the girl with her mother, her relatives, her church or her friends. And if none of them can or will help and the girl still adamantly insists on having and raising her baby, well now we're down to tough love. This is after all a democracy and if she wants to have the baby then that's absolutely her choice you step aside, wish her well and pray that God will help her because from this point forward it's out of your hands. nyt940830.0127

Little is known about the semantic, pragmatic and stylistic conditions under which the *'if'*-less expression of conditionality is preferred over the more prototypical *'if-then'*-form. As I see it, a treatment of these expressions would have to build on an existing theory of "conditionality"; the core of the latter is best developed by focusing on a narrow range of fairly well-understood data. I leave the extension to future work and will henceforth not deal with the examples discussed in this section.

Chapter 2

Probability

2.1 Introduction **35**

2.2 Sources of Uncertainty **35**

2.2.1 Non-determinism 35

2.2.2 Ignorance 36

2.2.3 Vagueness and ambiguity 37

2.2.4 Methodological remarks 39

2.3 Conditional probability: The Thesis **40**

2.3.1 Probabilistic inference 42

2.3.2 The Thesis as a pragmatic condition 43

2.3.3 The agenda 45

2.4 Implementation **47**

2.4.1 Probabilities of propositions 47

2.4.2 Probabilities of truth-functional sentences 48

2.5 Probabilities of conditionals **51**

2.5.1 The proposition *'if A then C'* 52

2.5.2 Intermediate values 56

2.6 Questions to be addressed **57**

2.1 Introduction

This chapter serves two purposes: To present arguments in favor of a probabilistic theory of conditionals, and to begin to implement such a theory. The discussion of the former relies on the extensive philosophical literature on probabilities and conditionals, in particular the work of Ernest Adams (1965; 1975; 1998). It is neither possible here nor the purpose of this work to provide an in-depth survey of that literature. Instead, I will select a number of arguments that seem particularly relevant.

The second question concerning implementation addresses certain difficulties which arise in combining a probabilistic theory of *validity* with a semantic theory of *truth*. I will outline the solution I adopt and point out a question it raises which has not been dealt with the literature. That question will be solved in Chapter 3.

2.2 Sources of Uncertainty

Probabilities are a way of encoding uncertainty. Uncertainty is modeled as the presence of multiple alternative possibilities, and probabilities are numbers attached to those possibilities in a certain systematic way. Everything else is open to interpretation, and each choice of interpretation is encumbered with a heavy load of metaphysical and epistemological implications. For now, I will mention those choice points only briefly, keeping the discussion general enough to stay clear of any commitments in those areas.

The first question to arise when dealing with uncertainty concerns its *source*. In dealing with natural-language sentences, there are (at least) three major sources of uncertainty:

2.2.1 Non-determinism

I am about to toss a coin. The coin has a history of coming up heads and tails with some ratio, and I have no way of knowing what the outcome will be this time.

The question is not settled until the toss is done. An examination of the physical properties of the coin or the relative frequencies of previous outcomes may give me good reasons to believe that it is more likely to come up heads than not, but the evidence is not conclusive.

There are two motivations for ascribing uncertainty of this kind to genuine non-determinism: one esoteric and one prosaic.[1] The first appeals to modern physics, the second to common linguistic usage. The first typically involves somewhat puzzling observations about sub-atomic processes, which are relevant to semantic theory only if it aims for a general and objective notion of truth. Not everyone considers it a linguist's job to cover that ground.[2]

The second motivation concerns everyday reasoning and language use, which treats some events as certain (e.g., that the coin will fall down) and others as unpredictable (e.g., that it will land heads.) A determinist may object to this, too, by claiming that the outcome is in fact determined by an array of circumstantial factors which we are merely not in a position to examine. However, such an objection would be "based on faith rather than on evidence" (Hausman, 1998, p. 185), and I see the burden of proof with the determinist. Coin tosses would not make for such a convenient textbook example, were it not for the shared intuition that some things cannot possibly be known.

For all practical purposes, we act as if some events were more likely than others, and as if complete certainty about the course of events were impossible to attain. That is all we need for a theory of the language we use.

2.2.2 Ignorance

The coin rolled off the table and is now lying on the floor. I cannot see whether it came up heads or tails, but I have no doubt that one of these possibilities is realized, and furthermore, I know that I can find out: I only have to take a closer look.

[1]These labels were used in a discussion of similar issues surrounding causation by Hausman (1998).

[2]Barry Schein (p.c.) does not.

Before I tossed the coin, I saw that one of its faces was slightly corroded and the other was clean. From a distance, the exposed face on the floor looks clean. It is more likely to be one face than the other. That, however, is not an objective fact. It is not the case that one of the faces has a higher objective chance of being exposed right now than the other. The respective degrees to which I believe the possibilities to hold, like the very fact that I entertain both as "live" possibilities, are solely due to incomplete information, not indeterminate facts.

For all practical purposes, we act as if some facts were unambiguously settled and could at least in principle be determined by inspection. Metaphysics aside, the contrast between these facts and the aforementioned indeterminate ones is real and linguistically relevant.

2.2.3 Vagueness and ambiguity

The third major source of uncertainty resides in language itself. Many linguistic expressions are by their very nature vague or ambiguous, but such vagueness or ambiguity cannot be located in the world outside of language.

2.2.3.1 Vagueness

The denotations of vague expressions are linguistic artifacts whose relationships to the objects and states of affairs they are ascribed to is mediated by observable quantities. The degree to which (2.1) is true is not directly measurable. I may possess precise and reliable information about John's height, yet be unsure whether he can be properly called '*tall*'. No amount of physical observation can remove that uncertainty.

(2.1) John is tall.

Thus it is misguided to think of such the degree to which John is tall as objective. But neither is it subjective in the sense of embodying a lack of information. It is the (language-internal) degree to which the predicate '*tall*' is applicable to the referent of '*John*', given his physical properties. As Russell (1923) put it,

> Vagueness and precision alike are characteristics which can only belong to a representation, of which language is an example. They have to do with the relation between a representation and that which it represents. Apart from representation, whether cognitive or mechanical, there can be no such thing as vagueness or precision; things are what they are, and there is an end of it. Nothing is more or less what it is, or to a certain extent possessed of the properties which it possesses. (p. 62)

2.2.3.2 Ambiguity

While vagueness is an integral part of the meanings of certain words, ambiguity arises in the map from strings to meanings.[3]

(2.2) a. I went to the bank.

b. I saw the boy with a telescope.

In general, the uncertainty in interpreting ambiguous expressions arises only on the receiving end, that is, in understanding. In a sentence like (2.2a), what the speaker conveys is not that she went to an amorphous place that somehow combines to different degrees the properties of a financial institution and the slope at the edge of a river. Nor does she convey that she went to a place that unmistakably is one or the other, but that she is in doubt as to what it was.

Likewise for syntactic ambiguity: (2.2b) typically is not used to convey that the speaker is uncertain as to whether she used a telescope to spot the boy or whether the boy she saw had a telescope.

[3]Put this way, the line may seem easier to draw in practice than it is. The criterion of the availability of continuous (for vagueness) versus discrete (for ambiguity) precisifications is commonly employed (Pinkal, 1985; Poesio, 1996), even though it does not always provide a clear-cut distinction.

2.2.4 Methodological remarks

2.2.4.1 Uncertainty

In what follows, I will not deal with all of the sources of uncertainty affecting the interpretation of language, but make the following choices and assumptions:

- Vagueness and ambiguity are left out of the discussion. In particular, the interpretation of simple sentences, negations and conjunctions is "crisp" in the sense that gradience in their denotations can only arise from objective or subjective uncertainty.
- Objective chance and subjective credence[4] follow the same probabilistic calculus and are not distinguished in the remainder of this chapter (they will be in the next chapter.)

In leaving out the purely linguistic factors of vagueness and ambiguity, I do not mean to diminish their importance. Nor do I mean to imply that they are less susceptible to a probabilistic analysis. I merely save them for future work.

2.2.4.2 Models

The central pursuit in this work is directed at *models* for the interpretation of language. It is important to be clear about what those models are supposed to be models *of*.

Models are abstract structures against which sentences in a given language are interpreted. Both truth and degree of support are defined at the interface between the language and the model. In scientific inquiry, the goal is to find the model that most faithfully matches reality, so that those sentences that are true in the model can be assumed to be true in the world. The match is evaluated by comparison against

[4] "Credence" is the term used by Lewis (1980) to refer to what Carnap (1945) had called "degree of confirmation."

empirical observations. In logic, on the other hand, the match with reality is of no relevance. What is studied is the relationship between a language and the class of its models, and relations between expressions in the language defined with reference to their model-theoretic denotations (such as entailment.) How "natural a world" or how "rational an agent" the model represents is an altogether different question.[5]

The use of model-theoretic methods in linguistic analysis follows the logical tradition: The question asked is *what if*, not *whether* an expression has a certain denotation. How well a model matches reality is as irrelevant as is the question as to how that match could be evaluated.

It is important to make this point explicit here because a probabilistic approach to natural-language semantics may face the sceptical question of where its numerical values "come from." Granted, in practice it is not clear whether the contribution of any of the various sources of uncertainty could be quantified with the precision that the customary use of real numbers would suggest. Nevertheless, in the sense in which the criticism is applicable, it rocks a boat that is shared by both probabilistic and truth-functional approaches. In all of what follows, very little use will be made of numbers, and where I do use them, they are merely intended as illustrations.

2.3 Conditional probability: The Thesis

The central tenet of the probabilistic treatment of conditionals I advocate is often simply referred to as "The Thesis."[6] It associates the value of a conditional *'If A, then C'* with the corresponding *conditional probability of C, given A*. A great deal of controversy surrounds the question as to what "value" is meant and what the association consists in.

The motivation of an analysis in terms of conditional probability is the Ramsey

[5] Cf. Chuaqui (1991) for a discussion of these two uses of models.

[6] Also referred to as "Stalnaker's Thesis" (after Stalnaker (1970); however, Stalnaker (1976) disavowed it); "Adams' Thesis" (after Adams (1965, 1975)); or the "Conditional Construal of Conditional Probability (CCCP)" (Hájek and Hall, 1994).

Test, discussed in Section 1.3.2.1. The intuitive connection becomes clear from the definition of the former. I will give a formal version in Section 2.4; for now, the following informal characterization will suffice. In a probabilistic setting, support of a sentence by a belief state is measured in degrees. Conditional probability can be paraphrased as in (2.3).

(2.3) The conditional probability of C, given A, is the degree to which C would be believed, were A observed (or learned) to be true.

Learning that A is true amounts to nothing else than adding the proposition *'that A'* to the prior set of beliefs. The addition of A may have repercussions throughout the belief set, raising the degrees of some beliefs, lowering those of others. The value assigned to C in the new belief set, *after* the (hypothetical) addition of A, is by this definition the degree of belief in the conditional *'if* A *then* C' *before* the addition. Thus (2.3) is merely a reformulation of the Ramsey Test in an enriched framework in which beliefs are quantified.

Ramsey (1929) himself made the connection between conditionals and conditional probability. Many authors have since noticed and endorsed the intuition (Jeffrey, 1964; Stalnaker, 1970, and others.). Van Fraassen (1976) noted:

> [T]he English statement of a conditional probability sounds exactly like that of the probability of a conditional. What is the probability that I throw a six if I throw an even number, if not the probability that: if I throw an even number, it will be a six? (pp. 272–273)

The Thesis is most widely discussed in the context of indicative conditionals. Although I believe that it is also applicable to counterfactuals (as do Adams, 1975; Edgington, 1995 and others,) I will not discuss that application for the time being. The remainder of this chapter is concerned with indicative (epistemic and predictive) conditionals.

2.3.1 Probabilistic inference

Aside from its intuitive appeal, the Thesis has the major advantage that in conjunction with a probabilistic theory of reasoning, it makes the right predictions about the validity of inference patterns involving conditionals. Most of the work in this area was carried out by Ernest Adams.

For detailed expositions of the theory, the reader is referred to the books mentioned above and the more recent Adams (1965, 1975, 1998). The gist of it can be summarized by the following central assumption:

(2.4) What is preserved in everyday reasoning is not truth, but (high) probability.

This principle is justified in Chapter 3 of Adams (1975) in terms of the desire of reasoning agents to be "right" in the long run with their probability estimates, assuming that they ground their actions on those estimates.

Adams devised a formal system of probability-preserving inference centered around the following notion of validity:

(2.5) An inference is probabilistically valid (*p-valid*) if and only if it is impossible for the premises to be highly likely while the conclusion is not highly likely.[7]

Under this assumption, classical validity (where conditionals are interpreted materially) is a necessary, but not a sufficient condition for p-validity (where the interpretation of conditionals follows the Thesis.) All p-valid inferences are also classically valid, but some classically valid ones are not p-valid. In particular, among the inferences with conditional conclusions, precisely the problematic ones are *not* p-valid:

1. Strengthening the antecedent:
   $A \to C$ does not p-entail $AB \to C$.

[7] Adams' formulation is as follows: A sentence φ follows from a set of sentences Γ if and only for all $\epsilon > 0$ there is a $\delta > 0$ such that for all conditional probability functions Pr such that $Pr(B) \geq 1 - \delta$ for all $B \in \Gamma$, $Pr(\varphi) \geq 1 - \epsilon$.

2. Contraposition:
   $A \to C$ does not p-entail $\overline{C} \to \overline{A}$.

3. Vacuous truth:
   $\overline{A}$ does not p-entail $A \to C$.

4. Hypothetical syllogism (aka transitivity):
   $A \to B$ and $B \to C$ do not p-entail $A \to C$.

   However: $A \to B$ and $AB \to C$ *do* p-entail $A \to C$.

These facts are discussed in considerable detail in Adams (1998). I will say no more here about these and other advantages of the theory. What it suggests is that a theory of conditionals in which the Thesis plays a central role has great promise as a model of the *reasoning* behind everyday uses of conditionals.

What remains to be shown is what exactly that role of the Thesis should be. In Adams' theory, conditionals have a somewhat peculiar status as far as the connection between truth and probability is concerned. As I will discuss in more detail below, unlike non-conditional sentences, a conditional probability cannot be interpreted as the probability that a proposition is true. As a consequence, non-conditional sentences do and conditionals do not have truth values, and the theory of inference is successful because it dispenses with the notion of truth altogether.

Moreover, conditional probabilities are not defined for sentences in which a conditional is embedded in a constituent of another conditional. As a consequence, Adams has no way of dealing with such sentences.

2.3.2 The Thesis as a pragmatic condition

While Adams' theory of inference successfully deals with the role of conditionals in reasoning, it is not accompanied with an account of their truth. This raises the question of its proper place in a semantic theory. One school of thought seeks to retain truth conditions in terms of the material conditional and appeals to the Thesis as a pragmatic condition on *use*. It is justified by the following reasoning:

(2.6) The sentence *'if* A *then* C*'* typically conveys that should it turn out that A is true, it will be safe to infer C. However, if $A \supset C$ is believed solely on the grounds that A is believed to be false, then subsequent evidence for A counts as evidence against $A \supset C$ and *not* for C. Therefore it would be misleading for a speaker in such a situation to inform a listener that *'if* A *then* C*'*.

Grice (1989, Chapter 4) suggests that conditionals carry a *conversational implicature*[8] which he calls the "Indirectness Condition" and paraphrases as follows: "There are non-truth-functional grounds for accepting $A \supset C$." Grice seeks to explain this condition in terms of his overall theory of conversation: $\neg A$ and C are both logically stronger than $A \supset C$, hence more informative; Grice's Maxim of Quantity demands that the logically stronger be chosen. The use of the conditional is therefore dispreferred, but sometimes necessary to balance off the Maxim of Quantity against that of Quality, in case the speaker does not have adequate grounds for either *'not* A*'* or *'*C*'*.

Jackson (1979, 1984, 1987) points out some counterexamples to Grice's argument and replaces the Indirectness Condition with an appeal to the Thesis, employing the conditional probability of C, given A, as a measure of the *assertibility* of *'if* A *then* C*'*.[9] Assertibility is to be distinguished from *truth*: He defines the latter in terms of the material conditional, whose probability is not equivalent to the corresponding conditional probability. As a consequence, the probability that a conditional is true may be high while its assertibility is low.

I agree with Jackson's critics who object that he reduces the notion of truth to a mere technicality of little use but with counterintuitive consequences. Appiah (1984) uses his theory to derive the wrong predictions about the assertibility of conditionals which appear as constituents of conditionals. To this, (Jackson, 1987, Appendix) replies that he does not mean for it to make *any* predictions in those cases, which he

[8] Jackson (1979) and, following him, Lewis in the Postscript to the reprint of Lewis (1976) in Jackson (1991), argue that it is a *conventional* implicature.

[9] Jackson (1984) uses the term "assertability"; Jackson (1987) changes the spelling to "assertibility" and reserves the earlier form for other pragmatic factors not related to conditional probability.

proceeds to declare either reducible to simpler conditionals or incomprehensible.

Nor does Jackson's definition of truth facilitate an explanation of the inference patterns discussed above. He maintains that such patterns do preserve truth, but not assertibility, and that in the case of conditionals it is the lack of the latter which tricks us into perceiving them as invalid (Jackson, 1987, p.47ff.). Thus in practical terms, one may believe that the premises of an argument are true and assertible and that its conclusion is true but not assertible.

Although Jackson gives truth values to conditionals, they play no useful role in his theory. Edgington (1995) seems right in concluding that "[w]e have yet to see any explanatory advantage of Jackson's theory" (p. 301). If it is really conditional probability that guides our intuitions about conditionals, then the theory of inference might as well be centered around it.

2.3.3 The agenda

The picture can be characterized as follows. The material conditional assigns truth values to conditionals but does not correctly predict when it is appropriate to infer or use them. The variably strict conditional, in its common application to indicative conditionals, makes the same wrong predictions. On the other hand, the Thesis combined with Adams' theory of inference corrects those predictions but does not deliver truth values. Jackson's theory sits somewhat uncomfortably between these extremes, assigning truth values of one kind and assertibility values of another.

The evidence that the Thesis governs our intuitions about inference and the use of conditionals is compelling. Jackson is right, however, in seeking to combine it with a theory of truth. One obvious reason for wanting truth values is coverage: Compounds and embeddings of conditionals are used and interpretable, but the Thesis does not define probabilities for them. However, any sentence with *truth* conditions can receive a probability.

The second reason concerns intuitions regarding what conditionals are about. If it is denied that they have truth values, then a speaker's belief in a conditional cannot be described as his belief that the conditional is true, and some other interpretation must explain what its probabilities is the probability *of*.

One may reply (as Appiah (1984) does) that conditionals are merely descriptions of epistemic states and inferential dispositions. This does not seem right, however: One can contest a conditional someone has asserted by calling it *false* without thereby denying its accuracy as a depiction of that person's beliefs; rather, what is contested is its correspondence with the *facts*, the assumption being that there is something "correct" to be believed about it. Consider, for instance, the sentences in (2.7a,b) from Edgington (1995):

(2.7) a. If that lump of sugar is placed in water, it will dissolve.
b. If that lump of granite is placed in water, it will dissolve.

Speakers do not hesitate to call (2.7a) *true* and (2.7b) *false*, regardless of whether the item in question is in fact placed in water or not. These are judgments of *objective* values which rational individuals should aim at approximating in their beliefs.

In what follows, I am going to propose a theory of objective values which obeys the Thesis and fulfills the desiderata just outlined. I do not call these values *truth values*, but they are objective nevertheless. They are assigned to conditionals with reference to the facts, not to beliefs. They also extend to compounds and embeddings of conditionals.

I do make an important semantic distinction, however, between predictive sentences like (2.7a,b) and epistemic ones like (2.8a,b):

(2.8) a. If that lump of sugar was placed in water, it dissolved.
b. If that lump of granite was placed in water, it dissolved.

While (2.7a,b) generally do have objective values, (2.8a,b) may not. What that amounts to will become clear in Chapter 4.

2.4 Implementation

The previous sections made a case for the suitability of a probabilistic analysis of conditionals in modeling reasoning. I proceed to provide the beginnings of a suitable formal semantics. This section fixes a few preliminary definitions, but stops short of providing a treatment of conditionals.

2.4.1 Probabilities of propositions

In modeling uncertainty in a possible-worlds framework, each open possibility is a set of worlds. Starting from a set W representing the set of all worlds that are at all "possible," a probability distribution is defined as a normalized *measure* on W, that is, a function from subsets of W to the real numbers between 0 and 1, inclusive:

Definition 1 (Probability model)
A probability model *is a structure* $\langle W, Pr\rangle$*, where* W *is a non-empty set of worlds and* Pr *is a* probability distribution *over* W*, i.e., a function* $Pr : \wp(W) \mapsto [0,1]$ *satisfying the following conditions for all* $X, Y \subseteq W$*:*

$$\begin{aligned} 0 \leq Pr(X) &\leq 1 \\ Pr(W) &= 1 \\ Pr(X \cup Y) &= Pr(X) + Pr(Y) \text{ if } X \text{ and } Y \text{ are disjoint} \end{aligned}$$[10]

Based on this, the notion of conditional probability of Y, given X, is defined as the "amount" of Y-worlds within the set of X-worlds whenever there are X-worlds:

[10] The third condition must hold in general for the limits of countable unions. That is important for the coherence of the underlying measure Pr, but of little concern in what follows since the language I deal with has only finite sentences.

Definition 2 (Conditional probability)
The conditional probability of Y, given *X for $X, Y \subseteq W$ is defined as follows:*

$$Pr(Y|X) = \begin{cases} \dfrac{Pr(X \cap Y)}{Pr(X)} & \textit{if } Pr(X) \neq 0 \\ \textit{undefined otherwise} \end{cases}$$

The conditional probability of Y, given X is not defined if X has zero probability. This is above all an artifact of the mathematical definition; one may make use or dispose of it in various ways. One way is by *stipulating* some value for the conditional probability in those cases, such as 1 (Adams, 1965; McGee, 1989), $Pr(Y)$ (Milne, 1997), or any other arbitrary value (Skyrms, 1988).[11] I do not adopt any of these stipulations and maintain instead that an undefined value is real and intuitively plausible in these cases and gives rise to pragmatic infelicity (cf. Chapter 4.)

2.4.2 Probabilities of truth-functional sentences

The last section defined probabilities of propositions. It is sentences, however, that are at issue here. I start with a set $\mathcal{A}$ of propositional *letters* and close the set under negation and conjunction (signified syntactically by a horizontal bar and concatenation, respectively.)

Definition 3 (Language $\mathcal{L}^1_{\mathcal{A}}$)
Given a set $\mathcal{A}$ of propositional letters, the language *$\mathcal{L}^1_{\mathcal{A}}$ is the smallest set containing $\mathcal{A}$ and closed as follows: If $\varphi, \psi \in \mathcal{L}^1_{\mathcal{A}}$, then $\overline{\varphi}, \varphi\psi \in \mathcal{L}^1_{\mathcal{A}}$.*

Note that $\overline{\varphi\psi}$ is the negation of $\varphi\psi$, whereas $\overline{\varphi}\,\overline{\psi}$ is the conjunction of $\overline{\varphi}$ and $\overline{\psi}$. Conjunction and the material conditional can be defined as usual: $\varphi \vee \psi$ and $\varphi \supset \psi$ as $\overline{\overline{\varphi}\,\overline{\psi}}$ and $\overline{\varphi\overline{\psi}}$, respectively. I refer to the material conditional using the

[11] The first of these options is "a version of the doctrine that a contradiction entails everything" (McGee, 1989), thus making an intimate connection to classical logic. Adams (1965) relied on it in building a theory in which classical validity has its place as an extreme case of probabilistic validity. But then again, the "doctrine" is responsible for the most counterintuitive aspects of classical logic.

symbol $\supset$. We already know, of course, that the material conditional is not a suitable connective to model the natural-language conditional. Below, I will use $\rightarrow$ for that "natural" conditional.

Sentences of the language are true at some worlds and false at others. The values 1 and 0, respectively, are assigned accordingly point-wise at individual worlds by a recursive *interpretation* function V. For each sentence A, $V(A)$ is the *characteristic function* of the proposition (i.e., set of worlds) in which A is true. The first version of V, V^1, is not yet defined for conditionals.

Definition 4 (Interpretation for $\mathcal{L}^1_{\mathcal{A}}$)
An interpretation *of the language $\mathcal{L}^1_{\mathcal{A}}$ in a probability model $\langle W, Pr\rangle$ is a function $V^1 : \mathcal{L}^1_{\mathcal{A}} \mapsto \{0,1\}^W$ satisfying the following conditions:*

$$\begin{aligned} \textit{For } A \in \mathcal{A} : V^1(A)(w) &\in \{0,1\} \\ V^1(\overline{\varphi})(w) &= 1 - V^1(\varphi)(w) \\ V^1(\varphi\psi)(w) &= V^1(\varphi)(w) \cdot V^1(\psi)(w) \end{aligned}$$

In statistical jargon, the set of possible worlds is the *sample space*, in which each world is an *outcome*. A proposition (a set of worlds) is an *event* in statistical usage. The characteristic function of a proposition is its *indicator function*. Indicator functions are a special kind of *random variables*.[12] Any function defined point-wise on the sample space is a random variable. By convention, its range is usually the set of real numbers or a subset thereof, such as $\{0,1\}$. Random variables may in general be continuous. I will ignore that case throughout for the sake of simplicity. There is no substantial difference between the discrete and the continuous case with respect to the matters dealt with here.[13] To generalize the calculations to the continuous case, summations are substituted with integrals.

[12]This use of the word "variable" to label a function has historical reasons and some potential to cause confusion. I follow common practice in using it; perhaps a term like "random quantity" (Chuaqui, 1991, p. 8) would be clearer.

[13]Moreover, using nonstandard notions, it is possible to simulate any stochastic process in continuous space and time by one defined on sets of unlimited size for which the usual summation is still appropriate. For an overview, cf. Nelson (1987); Chuaqui (1991).

I will often have occasion to refer to the set of worlds at which a sentence takes a particular value; for instance, the set $\{w \in W | V^1(A)(w) = 1\}$ is "the event that A is true." In the interest of readability, I abbreviate this expression as '$V^1(A) = 1$'; thus $Pr(\{w \in W | V^1(A)(w) = 1\})$ is alternatively written $Pr(V^1(A) = 1)$. Furthermore, I will refer to joint events using commas rather than intersection signs; thus $Pr(V^1(A) = 1, V^1(B) = 1)$ is written instead of $Pr(\{w \in W | V^1(A) = 1\} \cap \{w \in W | V^1(B) = 1\})$.

There is a straightforward relationship between the values of the characteristic function of a proposition and that proposition's probability: The latter is equal to the *expectation* of the former—i.e., the weighted average of its values, where the weights are the probabilities that the function takes those values. This notion, as well as the more general case of *conditional expectation*, will be important throughout.

Definition 5 (Expectation)
For random variables X, Y, expectation and conditional expectation are defined as follows:

$$E[X] = \sum_{x \in range(X)} x \cdot Pr(X = x)$$
$$E[X|Y = y] = \sum_{x \in range(X)} x \cdot Pr(X = x|Y = y)$$

I am mainly concerned with the probabilities of sentences, rather than propositions. These probabilities are defined in terms of the expectation of the assignment function:

Definition 6 (Probabilities of sentences)
Given a set of worlds W, the language $\mathcal{L}^1_{\mathcal{A}}$ with an assignment V^1, and a probability distribution Pr on W, a probability distribution P on the sentences in $\mathcal{L}^1_{\mathcal{A}}$ is defined as follows: For all $\varphi, \psi \in \mathcal{L}^1_{\mathcal{A}}$,

$$P(\varphi) = E[V^1(\varphi)]$$
$$P(\psi|\varphi) = E[V(\psi)|V(\varphi) = 1]$$

Since all sentences in $\mathcal{L}^1_{\mathcal{A}}$ denote functions with range $\{0,1\}$, the summation in Definition 5 is trivial and the relation between P and Pr is straightforward.

Fact 2
For all φ, ψ in $\mathcal{L}^1_{\mathcal{A}}$:

a. *$P(\varphi)$ is the probability that φ is true;*

b. *$P(\psi|\varphi)$ is the conditional probability that ψ is true, given that φ is true.*[14]

With these definitions, a number of facts about the probabilities of sentences follow as expected:

Fact 3
For all $\varphi, \psi \in \mathcal{L}^1_{\mathcal{A}}$:[15]

$$\begin{aligned}
P(\varphi) &= 1 \text{ if } \varphi \text{ is a tautology}\\
P(\varphi) &\leq P(\psi) \text{ if } \varphi \text{ entails } \psi\\
P(\overline{\varphi}) &= 1 - P(\varphi)\\
P(\varphi\psi) &= P(\varphi)P(\psi|\varphi)\\
P(\varphi \vee \psi) &= P(\varphi) + P(\psi) \text{ if } P(\varphi\psi) = 0\\
P(\varphi \supset \psi) &= 1 - P(\varphi) + P(\varphi\psi)
\end{aligned}$$

2.5 Probabilities of conditionals

The foregoing discussion was restricted to the language $\mathcal{L}^1_{\mathcal{A}}$ and its interpretation V^1, covering only the values of atomic sentences and truth-functional connectives thereof.

[14] *Proof.*

$$\begin{aligned}
P(\varphi) &= E[V^1(\varphi)]\\
&= 0 \cdot Pr(V^1(\varphi) = 0) + 1 \cdot Pr(V^1(\varphi) = 1)\\
&= Pr(V^1(\varphi) = 1)
\end{aligned}$$

$$\begin{aligned}
P(\psi|\varphi) &= E[V^1(\psi)|V^1(\varphi) = 1]\\
&= 0 \cdot Pr(V^1(\psi) = 0|V^1(\varphi) = 1) + 1 \cdot Pr(V^1(\psi) = 1|V^1(\varphi) = 1)\\
&= Pr(V(\psi) = 1|V(\varphi) = 1)
\end{aligned}$$

□

[15] Proof omitted.

The "natural" conditional is not part of the language so far. To set the stage for a discussion of what its values should be, I first extend the language to include it. The language $\mathcal{L}_{\mathcal{A}}$ in Definition 7 so far only includes "first-order" conditionals—ones which contain at most one '$\rightarrow$' as their main connective. This is sufficient to illustrate all of the main issues in this and the following chapter.

Definition 7 (Language $\mathcal{L}_{\mathcal{A}}$)
The language $\mathcal{L}_{\mathcal{A}}$ is the smallest set containing $\mathcal{L}^1_{\mathcal{A}}$ (cf. Definition 3, page 48) and such that for all $\varphi, \psi \in \mathcal{L}^1_{\mathcal{A}}, \varphi \rightarrow \psi \in \mathcal{L}_{\mathcal{A}}$.

Definition 8 provisionally extends the interpretation function to $\mathcal{L}_{\mathcal{A}}$, but the values of conditionals remain as yet undefined.

Definition 8 (Interpretation for $\mathcal{L}_{\mathcal{A}}$(partial))
For all $w \in W, \varphi \in \mathcal{L}^1_{\mathcal{A}} : V(\varphi)(w) = V^1(\varphi)(w)$.

The goal is to add a clause for sentences of the form $\varphi \rightarrow \psi$ to Definition 8 in such a way as to uphold the Thesis—the desideratum that the expectation of those values, $P(\varphi \rightarrow \psi)$, equal the conditional probability $P(\psi|\varphi)$.

This project faces a number of technical problems. In the remainder of this section I will discuss one of them and give a preview of the approach I adopt to address it.

2.5.1 The proposition '*if A then C*'

The denotation of an atomic sentence A is the characteristic function of a proposition, i.e., a set of worlds. For each world w, the value $V(A)(w)$ is fixed, and for any distribution Pr over the set W of worlds, the expectation $P(A)$ equals the probability $Pr(V(A) = 1)$ that A is true. The same holds for truth-functional compounds of atomic sentences.

Straightforwardly extending this strategy to conditionals, one would expect them to denote (characteristic functions of) propositions as well. The task then is to decide which worlds are in that set and which are not. It turns out, however, that there is no general strategy to do so.

The Thesis entails a number of properties which the sought-after proposition must satisfy, among them the following.[16]

Fact 4

Let φ, ψ be two truth-functional sentences. If $P(\varphi \to \psi) = P(\psi|\varphi)$ for all distributions Pr, then

a. *$\varphi \to \psi$ is consistent with φ;*
   i.e., $V(\varphi \to \psi)(w) = 1$ for some w such that $V(\varphi)(w) = 1$.

b. *$\varphi \to \psi$ is consistent with $\overline{\varphi}$;*
   i.e., $V(\varphi \to \psi)(w) = 1$ for some w such that $V(\varphi)(w) = 0$.

c. *$\varphi \to \psi$ is not entailed by $\overline{\varphi}$;*
   i.e., $V(\varphi \to \psi)(w) = 0$ for some w such that $V(\varphi)(w) = 0$.[17]

The task is to classify all worlds according to whether they *verify* or *falsify* the conditional. I already mentioned the generally accepted rule of Ramsey's that it is falsified whenever the material conditional is. Furthermore, I classify those worlds as verifying cases at which both antecedent and consequent are true.[18] Thus a conditional $\varphi \to \psi$ is false at worlds where φ is true and ψ is false, and true at worlds where both φ and ψ are true. In other words, at worlds where the antecedent is true, the value of the conditional is that of its consequent. This coincides with the value of the classical material conditional as well as the motivation behind the definition of conditional probability. It ensures that the probability of $\varphi \to \psi$ is that of ψ, given φ; that it is 1, given $\varphi\psi$; and that it is 0, given $\varphi\overline{\psi}$. Thus the first part of the definition is (2.9).

[16] A version of the following argument was first made by Carlstrom and Hill (1978) and is also found in Edgington (1995).

[17] *Proof.* By *reductio*. (a.) Suppose $\varphi \to \psi$ is not consistent with φ. Let $P(\varphi) = .8$ and $P(\varphi\psi) = .4$. Then $P(\psi|\varphi) = .5$ while $P(\varphi \to \psi) \le P(\overline{\varphi}) = .2$. (b.) Suppose $\varphi \to \psi$ is not consistent with $\overline{\varphi}$. Let $P(\varphi) = .2$ and $P(\varphi\psi) = .1$. Then $P(\psi|\varphi) = .5$ while $P(\varphi \to \psi) \le P(\varphi) = .2$. (c.) Suppose $\overline{\varphi}$ entails $\varphi \to \psi$. Let $P(\varphi) = .5$ and $P(\varphi\psi) = 1$. Then $P(\psi|\varphi) = .2$ while $P(\varphi \to \psi) \ge P(\overline{\varphi}) = .5$. □

[18] I will say more on the plausibility of this rule in Chapter 4.

(2.9) $$V(\varphi \to \psi)(w) = V(\psi)(w) \text{ if } V(\varphi)(w) = 1$$

(2.9) has the consequence that the conditional expectation of the values of the conditional over those worlds where the antecedent is true is the conditional probability:

Fact 5
Given (2.9), $P(\varphi \to \psi|\varphi) = P(\psi|\varphi)$.[19]

What should the values be at worlds where the antecedent is *false*? Fact 4 implies that they cannot be uniformly 1, as are those of the material conditional, or 0. The conditional must be true at some of those worlds and false at others. Jeffrey (1991) showed that in addition, the Thesis combined with (2.9) determines at "how many" of them the conditional must be true—formally, this is given by the conditional expectation of $V(\varphi \to \psi)$ over the set of worlds where the antecedent is false.

Fact 6
Given the Thesis and (2.9), $P(\varphi \to \psi|\overline{\varphi}) = P(\psi|\varphi)$.[20]

The joint import of Facts 5 and 6 has been noted before. It amounts to the following, first pointed out by van Fraassen (1976, p. 278):

Theorem 9
The Thesis and (2.9) entail that for all φ *and* ψ, φ *and* $\varphi \to \psi$ *are stochastically independent.*[21]

---

[19] *Proof.*
$$\begin{aligned} P(\varphi \to \psi|\varphi) &= E[V(\varphi \to \psi)|V(\varphi) = 1] \\ &= 1 \cdot Pr(V(\psi) = 1|V(\varphi) = 1) + 0 \cdot Pr(V(\psi) = 0|V(\varphi) = 1) \\ &= P(\psi|\varphi) \end{aligned}$$
□

[20] *Proof.*
$$\begin{aligned} P(\varphi \to \psi) = P(\psi|\varphi) &= P(\varphi \to \psi|\varphi)P(\varphi) + P(\varphi \to \psi|\overline{\varphi})P(\overline{\varphi}) \\ &= P(\psi|\varphi)P(\varphi) + P(\varphi \to \psi|\overline{\varphi})P(\overline{\varphi}) \\ P(\psi|\varphi)\frac{(1 - P(\varphi))}{P(\overline{\varphi})} &= P(\varphi \to \psi|\overline{\varphi}) \end{aligned}$$
□

Stalnaker (1976) states it as follows:

> [F]or any φ with probability between 0 and 1 exclusive, and for any ψ, the ratio of the measure of worlds in which $\varphi \to \psi$ is true to the measure of worlds in which it is false will be the same in the φ worlds as it is in the $\overline{\varphi}$ worlds. (p. 306; notation adjusted)

This is a peculiar condition[22] and there is no obvious way to enforce it with a general rule—general in the sense that it defines the truth value of the conditional at each world *once and for all*, regardless of the probability distribution Pr and in such a way that the Thesis is upheld when Pr is modified by updating with new information.

To see that there is a problem, consider the following simple example: A distribution Pr assigns probabilities strictly between 0 and 1 to $A, AC, A\overline{C}$ and B; also, B entails AC (i.e., the set of worlds at which B is true is a subset of those at which AC is true.) Thus the conditional probability of C, given A, is defined and strictly between 0 and 1. The Venn diagram in Figure 2.1 visualizes the scenario.

Theorem 9 shows that the conditional must be true in some but not all $\overline{A}$-worlds. Now consider the distribution Pr', which is like Pr except that it assigns zero probability to B (Pr' may have been obtained from Pr by learning that B is false.) The difference affects the probability of AC while that of $A\overline{C}$ is the same; hence the conditional probability of C, given A, is different as well. In order to uphold the Thesis, a corresponding difference would have to be found in the $\overline{A}$-worlds in order to "keep the balance" between A-worlds and $\overline{A}$-worlds. In consequence, there must be $\overline{A}$-worlds at which $A \to C$ is true under Pr but false under Pr'. But then $A \to C$ cannot denote a proposition.

[21] Proof omitted. The theorem is reminiscent of the *triviality results* of Lewis (1976, 1986c), but distinct. Lewis makes an additional assumption under which it follows that the *consequent* is stochastically independent of the antecedent. In Chapter 4 I show that this premise of Lewis' result is not satisfied by the current approach and argue that that is as it should be.

[22] Stalnaker continues: "But why should this be true? One does need some intuitive explanation."

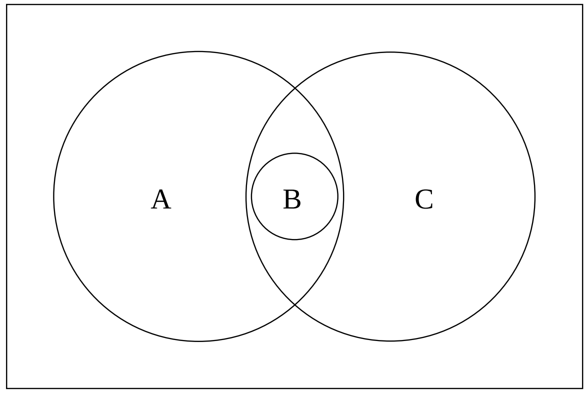


Figure 2.1: B entails AC

2.5.2 Intermediate values

Its headline notwithstanding, the last section ended in the conclusion that the Thesis is incompatible with the assumption that conditionals denote (characteristic functions of) propositions. I will now introduce an alternative, previously proposed by van Fraassen (1976) and Jeffrey (1991), which is not afflicted with the above problems.[23]

The upshot of the previous section can be summarized as follows: There is no general rule for assigning truth values to a conditional at worlds where the antecedent is false in such a way as to ensure that the Thesis hold for all probability distributions.

However, there *is* a general rule for assigning *values* at those worlds. As I mentioned in Section 2.4.2 on page 49, the characteristic functions denoted by sentences are indicator functions, a special kind of random variables. They are special in that their range is restricted to truth values, i.e., values in $\{0, 1\}$.[24] Random variables in general carry no restrictions on the values they take. In particular, they can fall between 0 and 1.

The basic idea of the random-variable approach is this: In general, some of the

[23] A related proposal is due to McGee (1989) and defended in terms of rational betting odds. It is similar in some respects but sufficiently different in others to be omitted in this section.

[24] These are the only values I am ready to call "truth values." Others are "values" simpliciter. Nothing hinges on this choice; I could have called them "intermediate truth values."

probability assigned to a conditional must be distributed over the worlds at which its antecedent is false. There is no general way of dividing those worlds into those at which the conditional is true and those at which it is false. But that is not required—instead, one can spread the probability evenly over all of those worlds. In general, this leads to the assignment of intermediate values (between 0 and 1, inclusive.)

Fact 6 on page 54 already established that the *expectation* of those values must be the conditional probability—which is also the expectation of the values the conditional takes where the antecedent is true. The most straightforward definition assigns this expectation uniformly to all non-antecedent worlds:

Definition 10 (Interpretation for $\mathcal{L}_{\mathcal{A}}$(complete))
Given an assignment function V^1 for $\mathcal{L}^1_{\mathcal{A}}$ in a model $\langle W, Pr\rangle$, an assignment V is defined for $\mathcal{L}_{\mathcal{A}}$as follows: For all $\varphi, \psi \in \mathcal{L}_{\mathcal{A}}$ and $w \in W$,

$$\text{If } \varphi \in \mathcal{L}^1_{\mathcal{A}} : V(\varphi)(w) = V^1(\varphi)(w)$$

$$V(\varphi \to \psi)(w) = \begin{cases} V(\psi)(w) \text{ if } V(\varphi)(w) = 1 \\ E[V(\varphi \to \psi)|V(\varphi) = 1] \text{ if } V(\varphi)(w) = 0 \end{cases}$$

Theorem 11
The interpretation function of Definition 10 enforces the Thesis for any φ, ψ, Pr for which the conditional probability is defined.[25]

2.6 Questions to be addressed

Theorem 11 shows that the values assigned according to Definition 10 conform to the Thesis. This is no reason to declare victory, however: On a closer look, it is not

[25] *Proof.*
$$\begin{aligned} P(\varphi \to \psi) =& E[V(\varphi \to \psi)] \\ =& E[V(\varphi \to \psi)|V(\varphi) = 1]Pr(V(\varphi) = 1) + \\ & E[V(\varphi \to \psi)|V(\varphi) = 1]Pr(V(\varphi) = 0) \\ =& E[V(\psi)|V(\varphi) = 1]\big(Pr(V(\varphi) = 1) + Pr(V(\varphi) = 0)\big) \\ =& Pr(V(\psi) = 1|V(\varphi) = 1) = P(\psi|\varphi) \end{aligned}$$
□

clear as yet whether any progress has been made towards a semantics of conditionals. Serious questions remain:

1. The assignment function of Definition 10 is a function not of two, but three parameters: a sentence, a world, and a probability distribution. It is the latter which determines the expectation assigned where the antecedent is false. How can this be reconciled with the basic motivation of truth assignments as depending only on the facts at the world of evaluation?

2. A related consequence of the dependence on the probability distribution is that different distributions induce different values at one and the same non-antecedent world. Is there any sense in which those values can be interpreted as the "objective values" given to the conditional at the world of interpretation? Should it not get just one value at each world?

3. Do all conditionals, or in any case all indicative conditionals, receive intermediate values as defined here?

Previous expositions (van Fraassen, 1976; Jeffrey, 1991) have not given these questions the attention they deserve and did not convince the audience that this "rather weird three-valued entity" (Edgington, 1995, p. 309) deserves a place in the theory of conditionals (Hájek and Hall, 1994, p. 100 voice strong reservations as well.)[26] Their respective motivations were of a rather abstract and formalistic nature.

Van Fraassen (1976) relates the values of non-antecedent worlds to the values of counterfactuals and borrows the basic idea from Stalnaker's semantics in terms of selection functions. Stalnaker's semantics is built around the notion that for each world at which the antecedent is false, there is a "nearest" world where it is true. However, van Fraassen suggests, that is too stringent a constraint. Instead, all that can be done is to choose one of the antecedent-worlds *at random*. The probability

[26]Indeed, Stalnaker and Jeffrey (1994, p. 34) explicitly refrain from proposing their elaboration on Jeffrey (1991) as an account of natural-language conditionals.

that the chosen antecedent-world is also a consequent-world is just the corresponding conditional probability.[27]

Jeffrey (1991) takes the idea of encoding sentences as random variables from de Finetti, starts by postulating that the Thesis holds and calculates what the value must be by an argument similar to the proof of Fact 6 on page 54.

Neither author gives a convincing answer to the above questions, however. It remains a mystery why the objective value of a conditional at a world should depend on, and change with, the probability distribution. In Chapter 3 I will resolve that question by appealing to objective *chance*, a time-variant distribution which depends only on the world (and time) of evaluation.

Furthermore, the strategy has been applied indiscriminately to all indicative conditionals; I will argue in Chapter 4 that it is appropriate for predictive conditionals but not for epistemic ones.

[27]This is a drastically simplifying paraphrase of the idea behind van Fraassen's "Stalnaker Bernoulli models," in which the probability in question is in fact defined as the expectation of an infinite sequence of trials.

Chapter 3

Time, truth, and knowledge

3.1 Introduction **61**
3.2 Time **61**
3.2.1 Intuitions 62
3.2.2 Historical necessity 64
3.3 Truth in time **67**
3.3.1 Temporal reference 67
3.3.2 Truth and settledness 70
3.3.3 Truth and future facts 71
3.3.4 Chance 72
3.3.5 Metaphysical remarks 73
3.4 Knowledge in time **74**
3.4.1 Partiality 74
3.4.2 Knowledge about history 75
3.4.3 Known and settled facts 77
3.4.4 Credence 78
3.4.5 Epistemological remarks 79
3.5 Probabilities of sentences **79**
3.6 Summary **81**

3.1 Introduction

The preceding chapter ended with a first implementation of a semantic interpretation which assigns objective values to sentences based on a world and a probability distribution. However, given that those values depend on the probability distribution, is not clear so far in what sense they can be called "objective."

I this chapter I address this question by suggesting that the probability distribution itself is best interpreted as terms of objective chance. I will develop models in which both objective chance and subjective probability ("credence") are encoded and determine the probabilistic values of sentences. I will also extend the interpretation function V to give time-sensitive values to sentences, but I will not discuss their probabilistic values (P in the preceding chapter) in any detail. This discussion is deferred to Chapter 4.

Section 3.2 discusses the intuitions about time that its representation is designed to accord with and makes the required modifications to the model. The remainder of the chapter is devoted the notions of *truth* and *chance* on the one hand (Section 3.3), *knowledge* and *belief* on the other (Section 3.4), and the relationship between them.

3.2 Time

In order to make the dependence on time explicit, the first task is to add a temporal dimension to the model. The view of time underlying my treatment is that of Thomason (1970), an "Ockhamist" notion of non-determinism in the terminology of Prior (1967). Burgess (1979) and Thomason (1970, 1984) also provide useful overviews. In Section 3.2.1 I will briefly discuss some intuitions motivating the formalism, which will be introduced in Section 3.2.2.

3.2.1 Intuitions

Simple sentences like *'It is Tuesday'* are sometimes true, sometimes false. Sentences like *'I have been in New York'* depend for their truth on past facts and may be true when uttered by a speaker who is not currently in New York.

To account for this time-dependence of truth, time is itself made a part of the model. In the approach I adopt, it is modeled as a set of points on which a strict linear order, the *earlier-than* relation, is imposed. The sets of facts that constitute worlds, structure-less under the previous approach, are then "stretched out" and lined up along the temporal dimension.

The present moment *now* is a moving point adding moments to the past and an intuitively significant divide between two kinds of facts: Past is "fixed" in a sense in which the future is not. Reichenbach (1956) summarized "the most obvious properties of time" in the following six "statements":

(3.1)
a. Time goes from the past to the future.
b. The present, which divides the past from the future, is now.
c. The past never comes back.
d. We cannot change the past, but we can change the future.
e. We can have records of the past, but not of the future.
f. The past is determined; the future is undetermined.

The asymmetry between past and future was noted by many other authors (Ramsey, 1929; Prior, 1967; Lewis, 1979, and elsewhere) and raises the question as to whether statements about the future have truth values at the time they are made.[1] The relevance of this question is evident for predictive conditionals, but it arises with respect to simple predictions as well. The sentences in (3.2a,b) are clearly felt to be

[1]A early statement of this question is due to Aristotle (On Interpretation 1:9): Any true (false) statement about past and present is necessarily true (false), but that is not the case for statements about the future: It is necessarily either true or false, but neither necessarily true nor necessarily false, that a sea battle will take place tomorrow.

either true or false. We may not know what those truth values are, but even lacking such knowledge, the question of their truth is *settled* one way or the other, and it cannot be different from what it in fact is.

(3.2) a. The coin landed heads.
b. The coin has landed heads.
c. The coin will land heads.

I assume that this is not so for (3.2c): It depends for its truth on future facts which, at the time it is asserted, are not yet actualized.

The intuition is embodied in models in which at each point in time, a single, fully specified past *branches* into multiple different futures. As the present moment "now" moves along the time-line, the tree is "pruned": At each time, exactly one of the states the world *could have* been in is added to the past as the one that it *is* in. All unactualized alternatives are discarded.

The multiple branches growing out of a common past represent uncertainty about the future. Given a fully specified past and present, this uncertainty is *objective* or *metaphysical*. To a possessor of information, it is the residual uncertainty which remains after all ignorance about past and present is removed.

This way of putting it is in fact still open to interpretation. There may be facts with the potential to distinguish worlds, but which could not possibly be known because the "language of knowledge" cannot distinguish between them. Two worlds which differ only with respect to such facts would be *distinct* without being *distinguishable*.

Viewed in this way, recognizing the asymmetry between past and future does not entail a commitment to the claim that the physical processes governing the world are truly non-deterministic. The required oppositions are still available if objective uncertainty is attributed to limitations of the expressive power of the language:

(3.3)

Non-determinism	Determinism
Identical pasts grow into *distinct* futures	*Indistinguishable* pasts grow into *distinguishable* futures

The choice between these options depends on one's metaphysical convictions. I am sympathetic to the non-deterministic view, but the models I use are suitable for both interpretations.

Total specifications of all facts past and present determine the *objective* values of sentences. Ordinary language use, of course, typically proceeds under partial ignorance of those facts. Such *epistemic* uncertainty, too, is relevant for a full account of conditionals: It determines their *subjective* values relative to information states. I will discuss both kinds of values and the relationship between them below.

3.2.2 Historical necessity

The models I use are variations on the "$T \times W$-frames" discussed by Thomason (1984), defined as in Definition 12. T is to be thought of as a set of moments in time, linearly ordered by the *earlier than* relation $<$. W is the set of worlds, as before.

Definition 12 ($T \times W$-frames (Thomason, 1984))
A $T \times W$ frame is a quadruple $\langle W, T, <, \approx \rangle$, where

1. *W and T are disjoint nonempty sets,*
2. *$<$ is a transitive relation on T which is also*
   (a) *irreflexive: $t \not< t$ for all $t \in T$; and*
   (b) *linear: For all $t, t' \in T$, either $t < t'$ or $t' < t$ or $t = t'$;*
3. *$\approx$ is a three-place relation in $T \times W \times W$, such that*
   (a) *for all t, $\approx_t$ is an equivalence relation; i.e., it is*
      i. *reflexive: For all $t \in T$ and $w \in W$, $w \approx_t w$;*

ii. *transitive: For all* $w, w', w'' \in T$, *if* $w \approx_t w'$ *and* $w' \approx_t w''$, *then* $w \approx_t w''$; *and*

iii. *symmetric: For all* $w, w' \in T$, *if* $w \approx_t w'$, *then* $w' \approx_t w$;

(b) *for all* $w, w' \in W$ *and* $t, t' \in T$, *if* $w \approx_t w'$ *and* $t' < t$, *then* $w \approx_{t'} w'$.

The worlds in W are thus "stretched out" in time and represented as sequences of momentary state descriptions, or "snapshots." As in the definitions in Chapter 2, they form the points of the probability space; Now they are to be thought of as complete *trajectories* through time, that is, functions from the temporal moments into the set of state descriptions. Where necessary to avoid confusion, I will refer to the worlds in this sense as "paths" or "world lines."

Each world evolves through time by shedding alternative futures. In the discussion below, it will be useful to reify those decreasing sequences of alternatives. I will call them *histories*, defined as follows:

Definition 13 (Histories in $T \times W$-frames)
For all $w \in W$ *and* $t \in T$, *let* $[w]_t^{\approx}$ *be the equivalence class* $\{w' | w \approx_t w'\}$ *of historical alternatives to* w *at* t. *A* history *is a function* $h : T \mapsto \wp(W)$ *such that for some world* $w \in W$, $h(t) = [w]_t^{\approx}$ *at all* $t \in T$.

Clause 3b in the definition of the relation $\approx_t$ of *being an alternative at* t is intended to ensure that the progressive loss of alternatives is not reversible. It is not possible for worlds to "become" historical alternatives at some point in time, nor is it possible for alternatives "not to have been" alternatives at earlier times. The effect of the definitions is this:

(3.4) Any world that is a member of a history at some time t is a member of that same history at all times up to t.

What would the evolution of a history through time in a $T \times W$-frame "look like"? Figure 3.1 gives an impression. Some of the world-lines are shown, many more are between the lines. Three histories in the sense of Definition 13 are represented

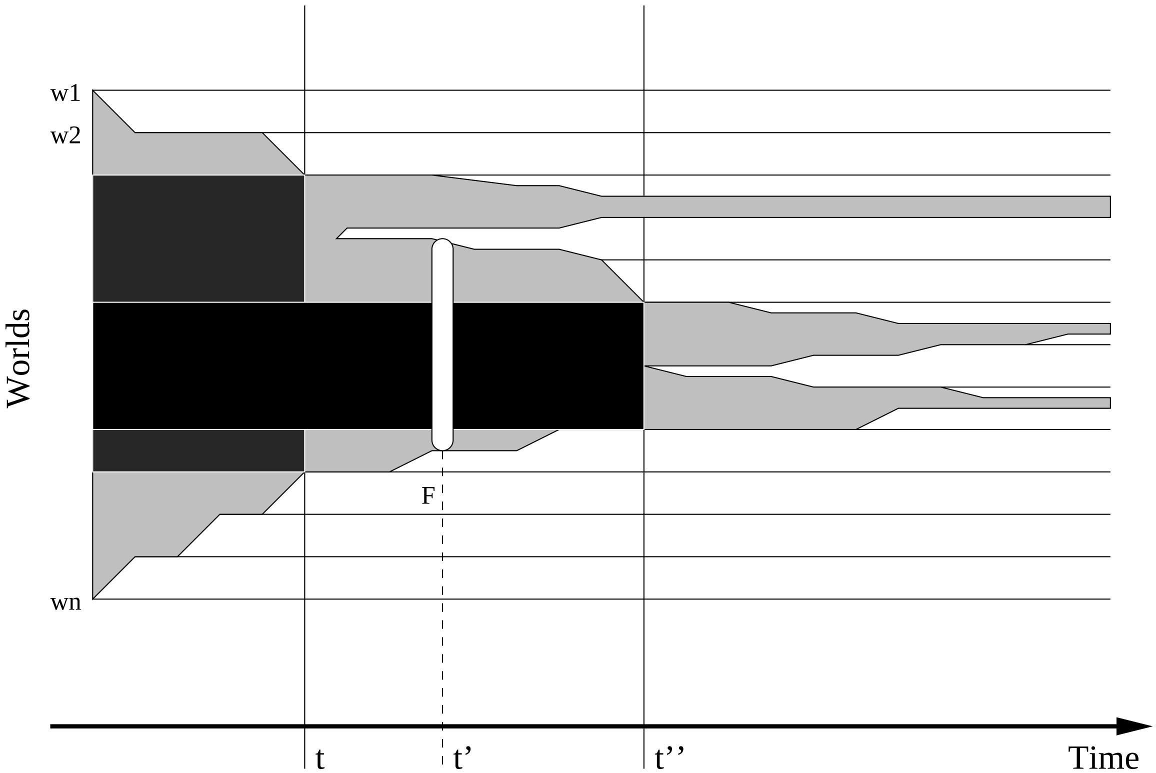


Figure 3.1: A $T \times W$ frame.

by narrowing grey areas. Each history includes ever fewer worlds as time progresses (from left to right). Different possible histories may agree up to some point (up to a short time after t in the picture, for instance,) but come apart at later times. The black areas enclose worlds that are indistinguishable up to some time: The worlds covered by the rectangle ending at time t, for instance, agree on all facts up to time t.

Thus historical alternatives to a world at a time can differ only with respect to facts at later times. This requirement is known as *Historical Necessity*.[2] Informally, it can be characterized as in (3.5). It serves to constrain the possible assignments of truth values to sentences, as spelled out in Definition 14 on page 68.

[2] The label has some potential to mislead: It does not mean that past and present *could not* have been otherwise; It means that past and present *cannot* have been otherwise.

(3.5) At all times t, the "snapshots" of world lines belonging to a common history at t are indistinguishable.

Given the monotonicity condition in (3.4) combined with (3.5), two worlds that both belong to a common history at some time t are indistinguishable at all times past and present from the perspective of t (though not necessarily at times later than t).

3.3 Truth in time

Recall that in Section 2.4.1 on page 47, the denotations of atomic sentences were taken to be characteristic functions of sets of worlds, that is, functions assigning truth values point-wise to worlds. For an interpretation which depends on time, truth values are assigned to atomic sentences not at worlds *simpliciter*, but at "snapshots," identified by their coordinates in the $T \times W$-space.

3.3.1 Temporal reference

The assignment function V can be encoded in different but interdefinable ways as a function from sentences to

(3.6) a. (characteristic functions of) time-world pairs;

b. functions from times to (characteristic functions of) sets of worlds: Each sentence at each time is assigned the set of worlds in which it is true at that time; or

c. functions from worlds to (characteristic functions of) sets of times: Each sentence at each world is assigned the set of times at which it is true in that world.

The difference is merely one of perspective; I shall adopt (3.6b). Thus for instance, the sentence in (3.7) denotes a function which, given a point t in time, returns a function which assigns 1 to all worlds in which John is in New York at time t, and 0 to all worlds in which he is not.

(3.7) John is in New York.

Formally, the definition runs as follows:

Definition 14 ($T \times W$ interpretation)
An interpretation *of the language $\mathcal{L}_\mathcal{A}$in a $T \times W$-frame is a function $V : \mathcal{L}_\mathcal{A} \mapsto \{0,1\}^{WT}$ from expressions in $\mathcal{L}_\mathcal{A}$to functions from members of T to (characteristic functions of) propositions, provided that for all atomic expressions in $\mathcal{L}_\mathcal{A}$, if $w \approx_t w'$ and $t_1 \leq t$, then $V(\varphi)(t_1)(w) = V(\varphi)(t_1)(w')$.*

As stated in the last section, in $T \times W$-frames worlds correspond to complete trajectories through time. Propositions are sets of such world-lines. The added temporal dimension affords a richer set of means by which to *identify* propositions, using sentences like (3.8a–c).

(3.8) a. John has been to New York.
b. Boys will be boys.
c. It was raining at 5am on July 2nd, 1999.
d. It was raining at 5am today.
e. It was raining.

To make the temporal reference involved in (3.8a–c) explicit in the formal representation, some expressive means must be added to the language. For (3.8a,b) this addition consists in the temporal operators **F**and **P**, interpreted as quantifiers over times. These operators are familiar since Prior (1967).[3]

Sentences (3.8c–e) all rely for their interpretation on reference to specific times, potentially different from that of evaluation: (3.8c) contains a temporal indexical,

[3]Prior also introduced the operators **H** and **G** for '*φ has always been / will always be the case*', respectively, defined as the duals of **P**and **F**. I will have no occasion to use them.

(3.8d) a function returning *'5am that day'* for each time of evaluation, and (3.8e) presupposes an implicit contextually given parameter. I will not be concerned modeling the way these reference times are fixed; instead, I assume that the sentences of the language carry their intended reference time "on their sleeves." In the interest of readability, I am going to commit the syntactic blunder of using elements of T, the set of times in the model, to label sentences, rather than distinct constants that are assigned to those times. No confusion is likely to arise from this.

The intended interpretation of sentences involving these expressions is given in (3.9a–c).

(3.9) a. $\mathbf{P}\varphi$: *'φ has been the case.'*

b. $\mathbf{F}\varphi$: *'φ will be the case.'*

c. $\varphi^{t'}$: *'φ is / was / will be the case at time t'.'*

In addition, I introduce an operator '**L**', also familiar from Prior (1967), whose interpretation I will discuss in the following section.

Definition 15 (Language $\mathcal{L}_{\mathcal{A}}^{T1}$)
The language $\mathcal{L}_{\mathcal{A}}^{T1}$ is the smallest set containing $\mathcal{A}$ and such that for all $\varphi, \psi \in \mathcal{L}_{\mathcal{A}}^{T1}$ and $t \in T$, φ^t, $\overline{\varphi}$, $\varphi\psi, \mathbf{P}\varphi, \mathbf{F}\varphi, \mathbf{L}\varphi \in \mathcal{L}_{\mathcal{A}}^{T1}$.

No modifications to the model are required to accommodate the new language. The assignment of truth values is now as in Definition 16.

Definition 16 (Interpretation for $\mathcal{L}_{\mathcal{A}}^{T1}$)
An interpretation V for the language $\mathcal{L}_{\mathcal{A}}$ in a model $\langle W, T, <, \approx\rangle$ (cf. Definition 14, page 68) is extended to an assignment V^T for $\mathcal{L}_{\mathcal{A}}^{T1}$ as follows:

For $\varphi \in \mathcal{L}_{\mathcal{A}}$: $V^T(\varphi)(t)(w) = V(\varphi)(t)(w)$

$$V^T(\varphi^{t'})(t)(w) = V^T(\varphi)(t')(w)$$

$$V^T(\mathbf{P}\varphi)(t)(w) = \begin{cases} 1 \text{ if for some } t' \text{ s.t. } t' < t,\ V^T(\varphi)(t')(w) = 1 \\ 0 \text{ otherwise} \end{cases}$$

$$V^T(\mathbf{F}\varphi)(t)(w) = \begin{cases} 1 \textit{ if for some } t' \textit{ s.t. } t < t', V^T(\varphi)(t')(w) = 1 \\ 0 \textit{ otherwise} \end{cases}$$

$V^T(\mathbf{L}\varphi)(t)(w)$ is defined in (3.10), page 70.

3.3.2 Truth and settledness

The last notion to be discussed here is that of truth in time with respect to *histories* rather than single worlds. Recall that the "indistinguishability" condition on histories in (3.5) above only mentions "snapshots": All the worlds that are members of the same history at some time t must be such that their snapshots at t are indistinguishable.

In addition, due to the monotonicity condition on histories, all those worlds must also be indistinguishable at all times prior to t. This ensures agreement among such worlds with respect to all sentences not containing the futurate operator **F**.

No such requirement is imposed for times later than t, however. Two worlds in the same history at time t may disagree with respect to future facts at later times, hence on the truth values of sentences involving reference to later times.

Thus the truth of a sentence at a world w and a time t no longer guarantees its truth at those worlds that are historical alternatives of w at t. Consequently, the *statement* that the sentence is true at w and t is not equivalent to the statement that it is true at all alternatives. The operator **L** distinguishes the latter from the former. The sentence $\mathbf{L}\varphi$ is true whenever φ is not merely true, but *settled*. (3.10) formalizes this notion and completes Definition 16.

(3.10) $$V^T(\mathbf{L}\varphi)(t)(w) = \begin{cases} 1 \text{ if for all } w' \in [w]_t^{\approx}, V^T(\varphi)(t)(w') = 1 \\ 0 \text{ otherwise} \end{cases}$$

To illustrate, consider a sentence φ which, for simplicity, does not contain temporal operators or indexicals. Given the conditions on histories, φ entails $\mathbf{L}\varphi$ and $\mathbf{P}\varphi$ entails $\mathbf{LP}\varphi$, but $\mathbf{F}\varphi$ does *not* entail $\mathbf{LF}\varphi$. On the other hand, $\mathbf{L}\varphi$, $\mathbf{LP}\varphi$ and $\mathbf{LF}\varphi$ entail φ,

$\mathbf{P}\varphi$ and $\mathbf{F}\varphi$, respectively. The distinction between truth and settledness is important in dealing with *predictions*, as the following section clarifies.

I will henceforth drop the superscripts from V^T and $\mathcal{L}_{\mathcal{A}}^{T1}$.

3.3.3 Truth and future facts

In the terms defined above, statements about the future may be *true* without being *settled*. It is important to distinguish these two notions because the choices I made above determine the kind of intuitions that are relevant in the treatment of conditionals below. In order to avoid confusion later on, I want to emphasize that it is settledness, not truth, that guides our judgments as to whether the *assertion* of a prediction is supported by the facts.[4]

Consider a time just before the toss of a fair coin. Only in some of the possible continuations of history does the coin come up heads, and it is impossible to tell whether the *actual* world will follow that course. Hence the use of (3.11) is not justified.

(3.11) The coin will come up heads.

However, that (3.11) cannot be asserted does not imply that it is false. The facts before the toss are irrelevant to its truth or falsehood, hence so is the justification (or lack thereof) lent to its assertion by those facts at that time.

Instead, the truth value of (3.11) depends on future facts *after* the toss: At those worlds where it later comes up heads (tails), it is true (false) *already*. Since all possible continuations are indistinguishable up to the time just before the toss, it is impossible to tell which is the actual one. Thus the value of (3.11) is not *settled*, hence I am not justified in asserting it even though it *may be* true.[5]

[4]This is closely related to Jackson's "assertibility." In Section 2.3.2 (page 43) I argued that his appeal to the notion is of little help because he does not define the assertibility of a conditional as the probability that it is true. I consider my use of it justified here because as I will show below, I do make that connection.

[5]This is the most prominent tenet of the "Ockhamist" view I adopt. It is not generally held;

3.3.4 Chance

At the end of Chapter 2, a probability distribution was defined as a function from sets of worlds to real numbers. No additional apparatus is needed to subject the models introduced in this chapter to a probabilistic interpretation. The probabilistic notion corresponding to *truth* is *chance*. The epistemic correlate, *knowledge* and *credence*, will be discussed in Section 3.4.

3.3.4.1 Models

By "chance" I mean, following Lewis (1980), the chance that a proposition is true. It is represented as a probability distribution over the worlds (now world-lines) in the model, similarly to the case of "timeless" models (Definition 1, page 47).

Definition 17 (Chance model)
A chance model *is a structure* $\langle W, T, <, \approx, Pr\rangle$ *where* $\langle W, T, <, \approx\rangle$ *is a* $T \times W$*-frame (cf. Definition 12, page 64) and* Pr *is a probability distribution over* W*.*

But chance does not remain constant over time. The chance that it will rain on the afternoon of a particular day changes with the facts up to the time in question. Once that time has passed, however, it is fixed: The chance that it *rained* at the time will forever remain 1 (if it did) or 0 (if it did not).[6]

In the model, this evolution through history depending on the facts corresponds to *conditionalization* on "what is the case." What is the case in a history h at a time t is the proposition $h(t)$. Hence the definition:

Definition 18 (Chance in history)
The chance of a proposition X *at a time* t *in a history* h *is* $Pr(X|h(t))$*.*

---

the opposing stance, according to which (3.11) is false in the situation described, is known as the "Peircean" or "Antactualist" view (Burgess, 1979, p. 575). The difference concerns the question which intuitions one chooses to regard as intuitions about truth.

[6]Cf. also the example of a man trapped in a labyrinth in Lewis (1980, p. 91).

3.3.5 Metaphysical remarks

A comment is needed here to address the metaphysical underpinnings of the models. In treating chance independently of the facts of the world, the definition admits models of any arbitrary course of history, including wildly counterintuitive ones. One can set up the probability distribution in such a way that a biased coin which came up heads consistently on any number of tosses throughout history has just a tiny chance of coming up heads on the next toss.

In terms of the flexibility and power of the model, that is as it should be: It must be possible to represent such scenarios. Nevertheless, one can hardly avoid wondering about *fit*, the question as to what would qualify as a "good" model of reality.

I realize that this question is important and needs to be addressed. It is not a linguistic question, however, and this is not the place to deal with it in any detail. I subscribe to the view that chance does not exist independently of, and indeed is determined by, the distribution of facts, and I suspect that it does not even exist independently of the observer—it may be no more than a conceptual construct, a product of our desire to see order where there is none.[7]

Given these tenets, one may find the models I use wanting in the following respects: If chance supervenes on facts, it is not clear how it could have existed *before* the facts. Moreover, since the facts are already embodied in the worlds, there should be no need to define probability independently. I do not deny that a metaphysical theory would have to address these questions, but devising such a theory is not an objective of this dissertation.

[7] I will not digress into a discussion of the philosophical literature on the subject. That literature is vast and varied; expositions of the view I adopt can be found, for instance, in Ramsey (1929); Carnap (1945); Lewis (1980, 1981a, 1994); Skyrms (1980, 1984); Hall (1994b); Hausman (1998). The task of *discovering* chance is akin to that of fitting a stochastic model to a stream of data. For a discussion of relevant notions in Artificial Intelligence, cf. Grünwald (1998).

3.4 Knowledge in time

In the preceding sections I developed a way of modeling uncertainty about the future from the perspective of an immutable past and present. In this context, in a history at a given time, only future facts whose status is not settled yet give rise to uncertainty. Any talk of this objective, metaphysical uncertainty "collapses" into talk of truth and falsehood when directed at settled facts.

This is not the only kind of uncertainty, however: Immutable facts need not be *known*. The truth values of (3.12a,b), evaluated at a time just after the coin was tossed, are settled, i.e., the same in all worlds which agree with the actual world up to the time of evaluation.

(3.12) a. The coin came up heads.

b. John lost his car.

c. If the coin came up heads, John lost his car.

John, who is blindfolded, is not aware of those truth values. He can venture a guess, however: Suppose he knows that the coin is biased and comes up heads 80 percent of the time. This knowledge has a bearing on the values he attributes to (3.12a,b) as well as to (3.12c).

Knowledge and the lack thereof play an important role in the interpretation of conditionals. In this section I outline a strategy of encoding incomplete information about history in the model.

3.4.1 Partiality

The progress of our world through time involves the elimination of alternative continuations. At each time t, new facts are added which may rule out some of the world lines that were "live" possibilities at times prior to t. The progress of history, then, really consists in *information gain*: Information about the way the world has been up to any given time t. This information is *objective* in the sense that it is not presumed to be entertained by an epistemic agent. It is information simpliciter.

The typically incomplete information entertained by an inhabitant of the world shares some properties with this objective information: The agent's uncertainty as to what the world is like can be modeled by spelling out the possibilities compatible with what he knows. Information gain proceeds by eliminating those possibilities that are incompatible with what is learned. Such updates are located in time, just like the accumulation of facts in objective histories. In these respects, the *epistemic history* of a reasoning agent is similar to the *objective history* he inhabits.

A similar distinction to that between *truth* and *chance* can be drawn between *knowledge* and *belief*. I will discuss the former in the next section and the latter in the section thereafter.

3.4.2 Knowledge about history

Where the notions of objective and epistemic history come apart is with respect to which facts can be subject to uncertainty: The epistemic agent may be unaware of past and present facts as well as future ones. While only one past is actual, the agent may not know which it is. This, too, is modeled as a set of possibilities: alternative pasts (and presents), representing ways of "filling in the blanks."

In Section 3.2.2 I discussed two requirements on $T \times W$-frames: (3.5) and (3.4), repeated here as (3.13) and (3.14):

(3.13) At all times t, the "snapshots" of world lines belonging to a common history at t are indistinguishable.

(3.14) Any world that is a member of a history at some time t is a member of that same history at all times up to t.

To model epistemic histories, I give up (3.13). An epistemic history at a given time may comprise worlds that *are* distinguishable. Those facts on which all of its members agree are *known*, the others are not. In this sense, the difference between *truth* and *settledness* that was observed with respect to future facts in the earlier case, is replicated as the difference between *truth* and *knowledge* with respect to past and present.

Definition 20 extends the $T \times W$-frames discussed earlier to encode these weaker conditions.

Definition 19 (Epistemic $T \times W$-frames)
An epistemic $T\times W$-frame *is a structure* $\langle W, T, <, \approx, \sim \rangle$ *where* $\langle W, T, <, \approx \rangle$ *is a* $T\times W$*-frame (cf. Definition 12, page 64) and* $\sim$ *is a three-place relation on* $T \times W \times W$ *such that*

1. *for all* t, $\sim_t$ *is an equivalence relation;*
2. *for all* $w, w', w'' \in W$ *and* $t \in T$, *if* $w \sim_t w'$ *and* $w \approx_t w''$, *then* $w \sim_t w''$; *and*
3. *for all* $w, w' \in W$ *and* $t, t' \in T$, *if* $w \sim_t w'$ *and* $t' < t$, *then* $w \sim_{t'} w'$.

For each time t, $\sim_t$ is the relation of being an *epistemic alternative* at t. Two worlds related in this way are equivalent in the sense that for all the agent knows, both may be the actual one.

Again similarly to the earlier definition of *histories*, we can define the *epistemic history* of the agent, the evolution of his knowledge over time, with reference to the $\sim$-relation.

Definition 20 (Epistemic histories)
For all $w \in W$ *and* $t \in T$, *let* $[w]_t^\sim$ *be the equivalence class* $\{w' | w \sim_t w'\}$ *of epistemic alternatives to* w *at* t. *An* epistemic history *is a function* $k : T \mapsto \wp(W)$ *such that for some world* $w \in W$, $k(t) = [w]_t^\sim$ *for all* t.

The import of Clause (2) of Definition 19 above is that $k(t)$ may not "cut across" possible histories $h(t)$. In particular, the agent is not capable of "presaging" the course of history by ruling out worlds that objectively are still open possibilities.

Accordingly, the definition of *knowledge* (Definition 21 below) will imply that nothing can be *known* unless and until it is *settled*.

Clause (3) above states that the accumulation of information is *monotonic*: Over time, the agent "zooms in" on a world, without ever forgetting anything. This is clearly an idealization, but a harmless one for present purposes.

Finally, notice that the definitions do not require the agent to consider the world which he inhabits an open possibility. He may be mislead into ruling it out.

3.4.3 Known and settled facts

The definition of what it is for a sentence φ to be *known* in an epistemic history parallels the one of objective settledness.

Definition 21 (Knowledge about history)
A sentence φ is known *at time t in an epistemic history k if and only if it is true (at t) at all worlds in $k(t)$.*

It follows at once that only *settled* facts can be *known*: By the definition of $\sim$, if a world is a member of $k(t)$, so are all its historical alternatives at t. The converse does not hold, however: Not everything that is *settled* need be *known*. That is as it should be. Consider (3.15).

(3.15) a. You struck the match.
b. The match lit.

I may know that the match did not light but be in doubt as to whether you struck it, even though the latter was settled earlier than the former.[8]

Furthermore, I may know that the question as to whether you struck the match is already settled without knowing what the answer is. As I pointed out in the introduction, the use of epistemic conditionals typically presupposes both the awareness

[8] I may also know *that* you struck the match without knowing *when* you did. The models can make this distinction, too, but I have no occasion to put them to that use.

that their consequent is settled and the lack of certainty as to which way it is settled. With respect to an epistemic history, this state of mind can be defined as follows:

Definition 22 (Presumption of settledness)
A sentence φ is presumed settled *in k at t if and only if for all $w, w' \in k(t)$, if $w \approx_t w'$, then $V(\varphi)(t)(w) = V(\varphi)(t)(w')$.*

A sentence is presumed settled if it is known that it is settled (i.e., if its objective value is constant for all ways of "filling in the blanks") without necessarily knowing what its value is.

3.4.4 Credence

According to Definition 22, an epistemic state $k(t)$ at a time t is compatible with a number of ways of "filling in the blanks." Each of those fully specified (up to t) histories assigns its own degree of chance assigned to a given proposition.

In addition, the agent may consider some of the ways of filling in the blanks more likely than others. These judgments represent the agent's *beliefs* about (as yet) unknown facts and are represented by a *credence* function Cr.[9] I assume that credence, like chance, can be modeled probabilistically. The two are not to be confused; however, the connection between them is intimate. I first give the definition of a *credence model.*

Definition 23 (Credence model)
A credence model *is a structure $\langle W, T, <, \approx, \sim, Pr, Cr\rangle$ where $\langle W, T, <, \approx, Pr\rangle$ is a chance model (cf. Definition 17, page 72), $\langle W, T, <, \approx, \sim\rangle$ is an epistemic $T \times W$-frame (cf. Definition 19, page 76) and Cr is a probability distribution over W.*

The final step is the definition of the relation between credence and chance. The degree to which an agent who finds himself in $k(t)$ believes a proposition X to be true is the *expectation* of the various values X takes in each of those fully specified histories:

[9]The use of the term goes back to Carnap (1945); see also Lewis (1980).

Definition 24 (Credence and chance)
The credence *given a proposition X, in k at t is the* expectation *of $Pr(X)$ conditional on the equivalence classes in $k(t)$ induced by $\approx_t$:*

$$Cr(X|k(t)) = \sum_{h(t) \subseteq k(t)} Pr(X|h(t)) \cdot Cr(h(t)|k(t))$$

3.4.5 Epistemological remarks

Definition 24 makes the assumption that the agent possessing the information in $k(t)$ would know how to calculate objective chance if only he knew all the facts up to t, i.e., which $h(t)$ is the actual one. This is an idealization which is most likely wrong (Thau, 1994; Hall, 1994b; Lewis, 1994). Assuming that chance is manifest in the overall distribution of facts, the actual chance is determined by *all* facts throughout history, including future ones. Since only an initial segment of history is settled at t, it is impossible at t to estimate chances correctly, even for an ideal agent who knows how to estimate chances from initial segments of history and who in addition knows all the facts up to t.

In terms of the model, this means that it cannot be assumed that Pr is available; instead, some approximation Pr' must stand in for Pr in Definition 24. I have no philosophical argument for my decision to use Pr nevertheless; I do believe, however, that the problem is not directly relevant to the linguistic matters I am concerned with.

3.5 Probabilities of sentences

In Section 3.3 I defined assignments for the truth-functional fragment the language enriched with tense operators. The assignment of probabilities is a straightforward extension of Definition 6 on page 50.

The distinction between chance and credence carries over to the probabilities of sentences. The chance assigned to a sentence is the chance that it is true, given the history up to the time of evaluation.

Definition 25 (Probability of sentences)

The chance *of a truth-functional sentence* φ *in history* h *at time* t *is*

$$\begin{aligned} P(\varphi|h(t)) &= E[V(\varphi)(t)|h(t)] \\ &= \sum_{x \in range(V(\varphi)(t))} x \cdot Pr(V(\varphi)(t) = x|h(t)) \end{aligned}$$

The values defined by P are the *objective* probabilities of sentences. Since for a sentences about settled facts (those at times up to and including t, the time of evaluation) the truth values at all worlds in $h(t)$ are uniformly either 0 or 1, they equal their expectation and chance is reduced to truth.

Similarly, the credence an agent lends to a sentence is merely the credence he attributes to its truth.

Definition 26 (Credence of sentences)

The credence given a sentence φ *in an epistemic history* k *at time* t *is*

$$C(\varphi|k(t)) = \sum_{x \in range(V(\varphi)(t))} x \cdot Cr(V(\varphi)(t) = x|k(t))$$

Recall that $k(t)$ is partitioned into multiple objective history slices $h(t)$, each representing a possible past and present (these slices are mutually exclusive, and they are jointly exhaustive since $k(t)$ cannot "cut across" any $h(t)$). It follows that for atomic and truth-functional sentences, credence equals the expectation of the various objective values the agent considers possible, weighed by the credence he gives each of the various possible histories bringing about those values.

Theorem 27

The credence of a truth-functional sentence φ *in an epistemic history is the expectation of its objective chance:*[10]

$$C(\varphi|k(t)) = \sum_{h(t) \subseteq k(t)} P(\varphi|h(t)) \cdot Cr(h(t)|k(t))$$

I will use both of the measures defined in this section as measures of the *support* given a sentence by a history (objective or epistemic) at a time. In the next chapter, I will argue that the same relation between credence and chance does *not* hold in general for conditionals.

3.6 Summary

This chapter developed the tools required for the semantic analysis of conditionals. All the ingredients—time, truth, chance, and credence—are in place. In the next chapter, I will proceed to discuss the probabilistic interpretation of various kinds of conditionals. It will turn out that all the ingredients are indeed needed to account for a variety of intuitions.

---

[10] *Proof.*

$$\begin{aligned} C(\varphi|k(t)) &= \sum_{x \in range(V(\varphi)(t))} x \cdot Cr(V(\varphi)(t) = x|k(t)) \\ &= \sum_{x \in range(V(\varphi)(t))} x \cdot \sum_{h(t) \subseteq k(t)} Pr(V(\varphi)(t) = x|h(t)) \cdot Cr(h(t)|k(t)) \\ &= \sum_{h(t) \subseteq k(t)} \sum_{x \in range(V(\varphi)(t))} x \cdot Pr(V(\varphi)(t) = x|h(t)) \cdot Cr(h(t)|k(t)) \end{aligned}$$

□

Chapter 4

The values of conditionals

4.1 Introduction **83**
4.2 Predictive conditionals **83**
4.2.1 The consequent 83
4.2.2 The antecedent 85
4.2.3 The conditional 87
4.3 Epistemic conditionals **88**
4.3.1 Objective values 88
4.3.2 Subjective values 89
4.3.3 Conditional credence and conditional chance 90
4.3.4 Subjective value assignment 91
4.3.5 Conclusion 93
4.4 Right-nested conditionals **94**
4.4.1 The Import-Export Principle 95
4.4.2 Constancy 100
4.4.3 Interpreting compounds of conditionals 101
4.4.4 The values at non-antecedent worlds 110
4.5 Summary **117**

4.1 Introduction

In Chapter 2 I introduced the basic approach to interpreting conditionals in such a way as to uphold the Thesis by assigning intermediate values at worlds where the antecedent is false. As it was presented there, it appeared stipulative and rather remote from the program of explicating the meaning and use of conditionals.

Chapter 2 discussed the model-theoretic background in which these apparent stipulations are sensible, in particular the dependence of the value of a conditional at a world on the probability distribution and, consequently, its ability to differ at one and the same world, depending on the distribution.

In this chapter I illustrate the application of the approach in the interpretation of various kinds of conditionals, the phenomena and distinctions it accounts for, as well as some of its limitations.

4.2 Predictive conditionals

Predictive conditionals are used to make conditional predictions. A typical example is (4.1).

(4.1) If you strike the match, it will light.

I will deal first with each constituent separately.

4.2.1 The consequent

Consider first the consequents of predictive conditionals, such as (4.2a). They typically contain the futurate modal *'will'*.

(4.2) a. The match will light.

b. $\mathbf{F}L$

Suppose (4.2a) is used at a time t to make a prediction. I represent (4.2a) as (4.2b). We can assume that the sentence is "about" a contextually salient future interval I within which the relevant facts become settled. Technically, the domain of the quantifier $\mathbf{F}$ is restricted to I; for simplicity, I leave this restriction implicit.

In assigning truth values to (4.2), the task is to divide the worlds in $h(t)$ into *verifying* and *falsifying* instances. The values at the worlds in a history h at time t, that is, at the members of $h(t)$, according to the rules of Chapter 3, are as in (4.3).

(4.3) $$V(\mathbf{F}L)(t)(w) = \begin{cases} 1 \text{ if for some } t' \text{ (in } I\text{) s.t. } t < t',\ V(L)(t')(w) = 1 \\ 0 \text{ otherwise} \end{cases}$$

That this rule is intuitively correct is best seen by looking at it as a guide in *discovering* the truth value of (4.2a). Since each world w in $h(t)$ is a fully specified trajectory through all of time, including times after t, it provides more information than the initial segment of its history up to t does. It answers questions about the future which are not yet settled at t. The value assigned at w and t is obtained by "looking ahead" to check how history will settle the matter, should it still contain w at the time in question.

This retroactive determination of the truth value by facts at later times does not imply that the actual value is (or was) *foreseeable* at t. Following the definition of truth in history in Chapter 3, all worlds in $h(t)$ are indistinguishable at all times up to and including t. Judgments made *at* t about the truth of the sentence can only be based on this initial segment of history, which gives no indication as to which continuation is the actual one.

Thus at t, what is determined about (4.2a) is not its truth, but only the *expectation* of its truth value, i.e., its probability $P(\mathbf{F}L|h(t))$, the measure of the *chance* (4.2a) has of being true. It is this probability that determines to what extent $h(t)$ *supports* (4.2a).

Definition 28 (Support)
The degree of support *of a sentence* φ *in a history* h *at time* t *is* $P(\varphi|h(t))$.

4.2.2 The antecedent

The antecedents of predictive conditionals are peculiar in that they refer to future times without the morphological marking they would have in isolation. I mentioned this property as an essential criterion in Dudman's classification of conditionals in Section 1.2.3 in the Introduction. I will now propose an explanation for this use.

The assignment of values to sentences at worlds involves identifying the set of those worlds at which the antecedent is true—for a conditional like (4.4a), the worlds at which (4.4b) is true. However, (4.4b) has a peculiar form; in isolation, (4.4c) tends to be used instead.

(4.4) a. If you strike the match, it will light.
b. You strike the match.
c. You will strike the match.
d. If you will strike the match, it will light.

Although the antecedent of (4.4a) clearly refers to future facts, when *'will'* is inserted, it alters the interpretation. I am not sure whether (4.4d) can be used sensibly under any circumstances; the closest thing to an interpretation would be one according to which whether or not the match lights depends not on whether you strike it, but on whether you (now) *intend* to do so.

Far-fetched as it is, this reading illustrates that (4.4d) is an epistemic conditional, used under the presumption that the antecedent is settled, and not a predictive one: Whether you intend to strike the match (4.4d) is settled, but whether you will actually do so (4.4a) is not. How does this difference come about?

I consider the fact that (4.4b) is not always infelicitous a promising lead. I mentioned in the introduction the observation that such "naked" present tenses, though uncommon with eventive predicates, can be used when the event referred to is "scheduled" (or planned by the "Almighty"). For instance, a film director may felicitously use (4.4b) in response to an actor's question *'What do I do in the next scene?'*

What is peculiar about these contexts is that the future facts referred to must be *settled*. The screenplay is written, the director is free to "look ahead," and most importantly, unlike in reality, where the actor may be prevented in various ways from playing the next scene according to the directions, there is no uncertainty as to whether the screenplay says or does not say that he strikes the match.

I will strain the analogy no further. The proposal is the following: The use of the naked present is licensed because the set of antecedent-worlds is identified by hypothetically "presaging" history and treating future facts as settled.

The truth definitions imply that the "naked present" is capable of identifying the desired class of worlds: Both of (4.5a,b) are claims about some contextually salient future time at which the relevant facts become settled.

(4.5) a. You will strike the match. ($\mathbf{F}S$)

b. You strike the match. ($S^{t'}$)

The truth value of (4.5a) is defined as in the last section, and that of (4.5b) as in (4.6).

(4.6) $$V(S^{t'})(t)(w) = V(S)(t')(w)$$

Thus assuming that the contextual restriction on the future operator in (4.6a) is such that (4.6a,b) are indeed "about" the same time, their truth values at t are equivalent.

Why is the use of the naked form in isolation restricted to certain contexts? In the previous section I suggested that the support of a sentence like (4.5a) by a history h at a time t is measured by $P(\mathbf{F}S|h(t))$, its chance of being true. I attribute this gradience to the presence of the modal *'will'*. In contrast, I suggest that (4.5b) is supported only if its truth is settled, i.e., if and only if $\mathbf{L}S$ is true in $h(t)$. Anything short of certainty is less than enough.

Finally, the form with *'will'* cannot be used in the antecedent because it is prone to induce an epistemic interpretation (with the presumption of settledness).

This proposal is admittedly somewhat tentative and calls for further research, including comparisons with uses of naked present tenses in other constructions. For now, I will assume that it is correct and that accordingly the interpretation of predictive conditionals does not call for a special treatment of the antecedent.

4.2.3 The conditional

Having discussed the constituents taken separately, we can now see how the values of the conditional are distributed over the various worlds in the history-slice $h(t)$.

(4.7) a. If you strike the match, it will light.

b. $S^{t'} \to \mathbf{F}L$

Recall that the probability of a conditional is defined as the conditional expectation of the values of its consequent, restricted to the set of worlds at which the antecedent is true. In light of the preceding sections, this formal definition can be related quite directly to the interpretation of the sentences: The set of worlds conditional upon which the expectation is measured is $\{w \in h(t)|V(S)(t)(w) = 1\}$, the set of of all those worlds at which *'You strike the match (at t')'* is true. The probability of (4.7) is the degree to which *'it will light'* is supported by that set. This is illustrated in (4.8).

(4.8) $$V(S^{t'} \to \mathbf{F}L)(t)(w) = \begin{cases} V(\mathbf{F}L)(t)(w) \text{ if } V(S^{t'})(t)(w) = 1 \\ E[V(S^{t'} \to \mathbf{F}L)(t)(w)|V(S^{t'})(t) = 1] \text{ otherwise} \end{cases}$$

Among the worlds at which the match is struck, those at which it lights are *verifying* instances for (4.8), and those at which it does not are *falsifying* instances. The values 1 and 0 are distributed accordingly.

The worlds at which the match is *not* struck do not qualify as verifying or falsifying instances; nevertheless, their history up to t justifies the assignment of the intermediate value at t. Recall that their history up to t is indistinguishable from all alternatives in $h(t)$, including those at which the match is struck. Thus intuitively, the intermediate value assigned at such a world is the expectation that the the match *would* have lit if it *had been* struck. I will further elaborate on this intuitive connection to counterfactuals below.

4.3 Epistemic conditionals

Epistemic conditionals are used under the presumption that the antecedent is settled (cf. Definition 22, page 78). The consequent may be either a prediction as in (4.9a) or an assertion as in (4.9b).

(4.9) a. If you struck the match, it will light.
b. If you struck the match, it lit.

The consequent does not require any more discussion. The status of the antecedent deserves attention because it is responsible for two peculiarities affecting epistemic conditionals: (i) their objective chance may be undefined, and (ii) as a consequence, their credence under incomplete information is not in general the expectation of their objective chance.

4.3.1 Objective values

In a history in which you struck the match, the chance that you did is 1 and the chance that you did not is 0. Conversely in a history in which you did not strike it. There are no non-trivial chances once the facts are settled.

The conditional probability is not defined if the condition has zero probability. Thus in those histories in which you did not strike the match, the conditional probability that it lit (or will light), given that you struck it, is undefined.

I take the semantic symptoms of an undefined value to be that the conditional in question cannot be judged true or false, and not even more or less likely. Indeed, that is the case with sentences like (4.10a) and markedly different from the case of predictive conditionals like (4.10b) used before the time in question. There it was sensible to say that even a world at which the antecedent is false lends some degree of support, expressed by an intermediate value, to the conditional.

(4.10) a. If you struck the match, it lit.

b. If you strike the match, it will light.

4.3.2 Subjective values

Some more discussion is required to see how the subjective value of an epistemic conditional is related to its objective values. A simple example will illustrate that its credence given an epistemic history cannot be the expectation of the corresponding conditional chances.

Consider an epistemic history k at time t, representing the beliefs of a hypothetical inhabitant of an objective history h^* who lacks knowledge of some facts but presumes that they are settled. $k(t)$ is partitioned into equivalence classes of worlds, each representing an objective history considered possible by the agent. Assume that $h^*(t)$, the agent's actual history slice, is among those considered possible, that the match was not struck in $h^*(t)$, and, for simplicity, that $k(t)$ contains one additional history $h'(t)$ in which the match was struck and lit.

In $h^*(t)$, the chance of the conditional is undefined; in $h'(t)$ it is 1. The expectation of the respective chances relative to $k(t)$ is therefore $1 \cdot Cr(h'(t)|k(t))$, the agent's subjective probability that h' is the actual history, which is less than 1. Notice that this is also the agent's subjective probability that the match was struck.

It is not, however, the agent's conditional credence that the match lit, given that it was struck: The match was struck only in $h'(t)$, which provides the domain for the conditionalization, and all worlds in $h'(t)$ are such that it lit; hence the conditional credence is 1.

It is intuitively clear which of these values correctly measures the degree to which the agent would believe (4.10): It should be 1, not the credence that the match was struck.

Definition 29 (Credence of conditionals)
The credence $C(\varphi \to \psi|k(t))$ given a conditional $\varphi \to \psi$ in an epistemic history k at time t is the conditional expectation

$$C(\varphi \to \psi|k(t)) = E[V(\varphi \to \psi)(t)|V(\varphi)(t) = 1, k(t)]$$

4.3.3 Conditional credence and conditional chance

The conditional expectation in Definition 29 is taken "globally" over the whole set of worlds at which the antecedent is true, not "locally" for each possible history slice $h(t)$ in $k(t)$. The value is best referred to as "conditional credence," rather than "expectation of conditional chance," and in fact, as the example showed, the two are not equivalent.

This raises a question for the assignment of truth values at individual worlds: If conditional credence is not the expectation of the objective values of conditionals, is it the expectation of some other variable—one which assigns *subjective* values at worlds at which the *objective* values are undefined? Would there be some intuitive rationale for why the values should come apart at one and the same world? I believe there is: A world at which the value of a conditional is undefined may nevertheless contribute to an agent's belief in its truth or falsehood.

As I mentioned in Section 3.4.1, page 74, epistemic histories share, at a "macroscopic scale," a number of properties with objective histories. Each objective history slice $h(t)$ is represented by the set of those historical alternatives which agree on everything that is *settled* at t. Put differently, "everything that is settled" describes a *partial world* (a fully specified initial segment of a history) which is characterized by the set of total worlds it may "grow into" and a probability distribution over those possibilities.

Much the same can be said about time-slices of epistemic histories. Each $k(t)$ consists of those worlds that agree on everything that is *known* at t. "Everything that is known" describes a partial world (an only partially specified initial segment of its history). This partial object has "blanks" in the past, giving rise to a "backward indeterminacy" which is in no way objective in nature, but nonetheless affects the perspective of the epistemic agent. Together with the agent's subjective probability distribution, this partial object determines the value of epistemic conditionals, hence so do the known parts of all the possible completions it may "grow into."

In the objective case, the assignment of values is fully determined by the time of evaluation because that point unambiguously separates the facts that are settled from those that are not. In the subjective case, the time of evaluation is *not* sufficient to determine the value: Different beliefs about the settled facts, held at the same time, give rise to different beliefs. Thus just as at one and the same world predictions about the future course of events are supported to varying degrees by different *settled segments* of its history, so too are beliefs about the further course of discovery supported to varying degrees by different *known parts* of that history. Hence there is no contradiction involved in assigning a subjective value where there is no corresponding objective value, for the two are supported by different collections of facts.

4.3.4 Subjective value assignment

Formally, what is called for is an assignment function for subjective values that is distinct from the objective assignment V. The latter depends only on w and t, for these two parameters determine the unique $h(t)$ with respect to which values are assigned. However, w may, at t, belong to different epistemic histories, $k(t)$ and $k'(t)$; which of these it belongs to determines which of the facts of w enter the calculation of the intermediate values of conditionals.

Definition 30 (Epistemic interpretation)
Given an epistemic history k, the interpretation V for $\mathcal{L}_{\mathcal{A}}^{T}$ is extended to an epistemic interpretation V_k *as follows:*

$$\text{For } \varphi \in \mathcal{L}_{\mathcal{A}}^{T1}\text{: } V_k(\varphi)(t)(w) = V(\varphi)(t)(w)$$

$$V_k(\varphi \to \psi)(t)(w) = \begin{cases} V(\varphi \to \psi)(t)(w) \text{ if defined;} \\ E[V(\varphi \to \psi)(t) | V(\varphi \to \psi)(t) \text{ defined}, k(t)] \\ \qquad \text{otherwise.} \end{cases}$$

Since the objective value of a conditional is undefined at just those worlds at which the antecedent is true, it follows that the assignment of Definition 30 yields the right values to comply with Definition 29.

Fact 7
The credence $C(\varphi \to \psi | k(t))$ is the expectation of the epistemic values $V_k(\varphi \to \psi)(t)$.[1]

The two assignments V and V_k come apart only at worlds at which the conditional chance is not defined. This is only the case for antecedents that are settled, hence for epistemic conditionals, if we assume that all histories give *some* non-zero chance, however slim, to *futurate* antecedents and that therefore "forward-looking" conditional chance is always defined.[2]

To conclude this section, I want to point out that the assignment of distinct epistemic values accounts for examples of the "Sly Pete" kind (Gibbard, 1981). It is possible for two agents in different epistemic states too believe with certainty $A \to B$ and $A \to \overline{B}$, respectively, both relying solely on true collections of facts and both being completely rational. Edgington (1995) discusses this example at some length and shows that it only arises when it is settled that A is false and that fact is not known to either agent. Gibbard had used this case to argue that certain conditionals never have truth values. In fact, that is the wrong generalization: surely (4.11) is *falsified* if it turns out that the match was struck and did not light.

[1] *Proof.* $V(\varphi \to \psi)(t)(w)$ is defined if and only if $V(\varphi)(t)(w) = 1$. □

[2] Technically, "impossible" events may not be eradicated, but reduced to infinitesimal probability, so that conditional probabilities are still defined (Lewis, 1980; Skyrms, 1980; McGee, 1994).

(4.11) If you struck the match, it lit.

Instead, whether a conditional has an objective value depends on whether the corresponding objective chance is defined (For more discussion, cf. Edgington (1995), Section 8.2, pages 293–296).

4.3.5 Conclusion

Epistemic conditionals are peculiar in that their credence is not equivalent to the expectation of their conditional chance. In the last section I proposed an explanation for this observation which appeals to the fact that even though all facts of the world up to the evaluation time t are *settled*, only *known* facts determine the value of an epistemic conditionals at that world. This justifies the assignment of a subjective value to a conditional at a world where its antecedent is false by an agent who does not know that it is false, similarly to the assignment of an objective value at worlds where the antecedent is false but its falsehood is not settled. In each case, the value is calculated as the expectation over *alternatives*—historical alternatives in the objective case and epistemic alternatives in the epistemic case.

Stepping back, the results are quite similar: The values of both kinds of conditionals are true where both antecedent and consequent are true, false where the antecedent is true and the consequent is false, and where the antecedent is false, intermediate (if its falsehood is not known or settled, respectively) or undefined (if its falsehood is known or settled).

In what follows, the distinction between epistemic and objective values will not be crucial. The remainder of this chapter addresses a problem that affects both kinds, and undefined values will not figure in either the discussion or the examples. I will therefore indiscriminately refer to generic three-valued assignments. The discussion will cast doubt on the definitions of the intermediate value I have employed so far.

4.4 Right-nested conditionals

The scope of the discussion so far was limited to cases of "first-order" conditionals, sentences in which the arrow is the main connective and combines only truth-functional compounds of atomic sentences on either side. Natural language is not so simple, however: Conditionals may themselves embed other conditionals. Examples (4.12) and (4.13), perfectly well-formed sentences of English, illustrate that such constructions are not uncommon.

(4.12) a. "I wouldn't be surprised if Aim's management group emerged as the leader if these two firms agree to combine their businesses," said Richard Chimberg, editor of Investment Management Weekly, a weekly newsletter published in Boston.

nyt960926.0322

$(A \to (B \to C))$

(4.13) a. "We'll be fine if we just play our responsibilities and if something breaks down, pick up and be ready to go to the next play.

nyt960906.0664

$(A \to B) \wedge (C \to D)$

The philosophical literature on conditionals features its own share of standard "armchair" examples, some of which I will discuss below.

An advantage of the intermediate-value approach in terms of random variables over the traditional probabilistic calculus is that while conditional probabilities cannot be nested (an expression of the form '$P(C|A|B)$' is meaningless), assignments of intermediate values extend naturally to such sentences. In this chapter, however, I will limit the discussion to examples of the kind illustrated in (4.12) and (4.13).[3]

[3]The reason for this choice is that these sentences can be interpreted without complicating either the model or the interpretation function. An account of conditionals with conditional antecedents requires the would require the "Stalnaker Bernoulli"-models of van Fraassen (1976); Stalnaker and Jeffrey (1994), which would complicate the discussion very substantially.

In this section I will discuss two alternative ways of defining the values of right-nested conditionals, one (the "Import-Export Principle") employed by McGee (1989) the other (which I dub "Constancy") assumed by Jeffrey (1991). The McGee conditional makes what would appear to be intuitively correct predictions, which is why Stalnaker and Jeffrey (1994) credit it with the potential to be "a better representation of right-nested conditionals in natural language." (p. 40) On the other hand, the Jeffrey conditional is a mathematically more elegant solution which, moreover, can by stipulation be made to express the McGee conditional. Which of the two is correct?

4.4.1 The Import-Export Principle

The standard probabilistic calculus does not define values for right-nested conditionals like the one in (4.12). The Thesis would require both equalities in (4.14) and (4.15) to hold.

$$P(\varphi \to \psi) = P(\psi|\varphi) \tag{4.14}$$

$$P(\chi \to (\varphi \to \psi)) = P(\varphi \to \psi|\chi) \tag{4.15}$$

The problem is that the probabilistic calculus does not provide a way to substitute (4.14) in (4.15). As a workaround, one may take recourse to the "Import-Export Principle" of Definition 31 below and define "$P(\varphi \to \psi|\chi)$ to equal $P(\psi|\varphi\chi)$.

Definition 31 (Import-export)
For all distributions P and sentences φ, χ, ψ: $P(\varphi \to \psi|\chi) = P(\psi|\varphi\chi)$ whenever $P(\varphi\chi) \neq 0$.

There are good arguments against such a simplification, however. Technically, the Import-Export Principle leads to Lewis' *triviality results*, which I will briefly discuss in the next section. The triviality results do not speak decisively against the Import-Export Principle as such, however, only against Definition 31; McGee (1989) found a way to enforce it while avoiding the triviality results. The real argument against the validity of the Principle therefore appeals to its intuitive invalidity, which I will illustrate in the following section.

4.4.1.1 Triviality

A host of theorems summarily known as *triviality results* (Lewis, 1976, 1986c; Hall, 1994a) provide strong evidence that the denotations of conditionals cannot be propositions in the usual sense. The first result of Lewis (1976, 1986c) is replicated in Theorem 32.

Theorem 32 (Lewis (1976))
Let P be a probability distribution observing the Import-Export Principle. Then for all φ, ψ such that $P(\varphi \to \psi) = P(\psi|\varphi)$, φ and ψ are stochastically independent.[4]

Independence of φ and ψ means that learning φ will not affect the agent's belief about ψ—surely an absurd requirement, given that conditionals are typically used to convey that their constituents are *not* independent in this sense.

More could be said about this result and later, more powerful ones subsuming Theorem 32. I will not digress into that history here because the intermediate-value approach, in giving up the requirement that conditionals denote a proposition in the usual sense, is not vulnerable to them. (Hájek and Hall (1994) and Hájek (1994) provide an excellent overview). Here I will merely outline why the Import-Export Principle does not apply in the present formalism, why it *should not* hold, and why it nevertheless appears plausible in many cases.

4.4.1.2 Two readings of the Import-Export Principle

The Import-Export Principle involves an equivocation between two ways of reading the bar '|' in expressions like (4.16), given in (4.16a) and (4.16b):

[4] *Proof.*

$$\begin{aligned} P(\psi|\varphi) &= P(\varphi \to \psi) && \text{(Thesis)} \\ &= P(\varphi \to \psi|\psi)P(\psi) + P(\varphi \to \psi|\overline{\psi})P(\overline{\psi}) \\ &= P(\psi|\varphi\psi)P(\psi) + P(\psi|\varphi\overline{\psi})P(\overline{\psi}) && \text{(Import-Export)} \\ &= 1 \cdot P(\psi) + 0 \cdot P(\overline{\psi}) = P(\psi) \end{aligned}$$

□

(4.16) $P(\varphi \rightarrow \psi|\psi)$ is ...

a. the probability $\varphi \rightarrow \psi$ will (would) have if ψ turns out (turned out) true.

b. the expectation of the values of $\varphi \rightarrow \psi$ at those worlds at which ψ is true.

In Chapter 2 I used (4.16a) as an informal paraphrase of the concept of conditional probability. On the other hand, my definition of the probability of a sentence as the expectation of its values follows (4.16b).

Reading (4.16a) makes the Import-Export Principle highly plausible. What Lewis showed is that linking it to a possible-worlds semantics along the lines of (4.16b) leads to absurdity under the implicit assumption that the conditional denotes a proposition, in the sense that its values at individual worlds are (i) either 0 or 1 and (ii) constant. Under that assumption, the probability in (4.16) can only be 1, as the intuition behind (4.16a) dictates, if the conditional is true at all worlds at which ψ is. Likewise, the conditional must be false at all worlds at which ψ is false, since $P(\varphi \rightarrow \psi|\overline{\psi})$ should be 0. Hence the conclusion in the proof of Theorem 32 that $P(\varphi \rightarrow \psi) = P(\psi)$, which by the Thesis implies that $P(\psi|\varphi) = P(\psi)$.

In contrast, the two readings come apart in the current formalism. The definition of the probabilities of sentences does not validate the Import-Export Principle on reading (4.16b): The value of $\varphi \rightarrow \psi$ at an $\overline{\varphi}\psi$-world is in general intermediate (although it *may* be 1) therefore the expectation over all ψ-worlds does not in general add up to 1.

Recall, on the other hand, that the value at a $\overline{\varphi}\psi$-world is obtained from the overall probability of ψ, given φ. This is *guaranteed* to be 1 (unless it is undefined) whenever the probability of ψ is 1—which is the case if ψ is *settled.*[5] In other words, then, Reading (4.16a) is observed by the intermediate-value approach and (4.16b) is not. The difference between the two can be brought out by the paraphrases in (4.17a,b):

(4.17) $P(\varphi \rightarrow \psi|\psi)$ is ...

[5]Ignoring the mathematical possibility of a non-empty set of $\overline{\psi}$-world with probability 0.

a. the probability of $\varphi \to \psi$ given that ψ is *settled*. (must be 1)

b. the probability of $\varphi \to \psi$ given that ψ is *true*. (need not be 1)

A little reflection shows that the current approach is right in allowing for (4.17b) to fall short of 1. Here is an example.

Suppose at a time t, before the match is struck, the chance of it lighting if struck is .7. There is a good chance that you will strike it, but you may not: I may toss it into the fire instead, where it will certainly light without being struck. According to the Principle, the probability at time t of the conditional in (4.18a), assuming that the match lights, i.e., $P(S \to L|L)$, should be $P(L|SL) = 1$, *regardless* of whether it is struck at those worlds or not.

(4.18) a. If you strike it, it will light.

b. If you struck it, it lit.

c. If you had struck it, it would have lit.

That is not, however, the correct *prior* value (i.e., value at t): Rather, it is the *posterior* value of (4.18b), correctly assigned at a later time t' at which it is settled that the match lit. In contrast, at t' the value that (4.18a) *had* at time t would best be elicited by (4.18c). But of course, at worlds where the match lit without having been struck, that value should not be 1, but .7.

Intuitions about right-nested conditionals bring out the difference even more vividly. According to the Import-Export Principle, the probability of (4.19a) should be (4.19b):

(4.19) a. If it lights, then if you strike it, it will light.

b. $P(L \to (S \to L)) = P(S \to L|L) = P(L|SL) = 1$

Given the scenario, it is not at all clear that (4.19a) is true. I may consider it highly likely that the match will light because I am strongly inclined to toss it into the fire. Then "almost all the worlds," figuratively speaking, which I consider open possibilities are such that the match lights. According to the Principle, then, I should consider (4.18a), the consequent of (4.19a), at least that likely. But I don't: Surely the match will light if tossed into the fire, but that does not ensure that it will also light if struck.

4.4.1.3 Quasi-Import-Export

These arguments show that contrary to the Import-Export Principle, the first equality in (4.19b) is not valid. What about the second one? This question is important because (McGee, 1989) reaches the latter without going through the former, hence not falling prey to the triviality result, by defining $P(L \to (S \to L))$ directly as $P(LS \to L)$ (and so on, recursively for deeper right-embeddings).

How precisely this move allows McGee, p. 178 manages to "wiggle out of the trivialization argument" (Stalnaker, 1991) is not important here; what is important is that it is to no avail: The latter is intuitively not equivalent to the former either. Given the scenario, (4.20a) is felt to be trivially true, unlike (4.19a).

(4.20) If you strike it and it lights, it will light.

But is the Import-Export Principle not eminently plausible, the counterarguments just presented notwithstanding? Indeed, it often is, so much so that it is often assumed without much discussion.

However, the examples discussed here show that the inference is not validated by the *form* of the sentences, hence it is not a logical relation and the substitution of one for the other is not generally safe. Something else must be at work in the cases where it does seem plausible. Those include the standard version of the "wet-match" sentence in (4.21a), McGee's example in (4.21b), and others.

(4.21) a. If the match is wet, then if you strike it, it will light.

b. If you are asked to submit to a "voluntary" urine test, then if you refuse, you'll be under suspicion.

The question of what it is that supports the inference in these examples is at the center of this section.

4.4.2 Constancy

The Import-Export Principle is embraced by McGee (1989) and rejected by Jeffrey (1991), who writes:

> I take it that if $A \to C$ is to be an *indicative* conditional it must have the same value at all worlds w where A is false ... (p. 172)

This assumption is implicit in the definition of assignments to simple conditionals (Def. 10, p. 57) which is adopted from Jeffrey. I will henceforth refer to it as the *Constancy* condition. It states that the conditional expectation of the values assigned at the non-antecedent worlds is independent of any "third" facts:

Definition 33 (Constancy)
For all assignment functions V, sentences φ, ψ, χ and real numbers x:

$$E[V(\varphi \to \psi)|V(\varphi) = 0, V(\chi) = x] = E[V(\varphi \to \psi)|V(\varphi) = 1]$$

The incompatibility of this condition with the Import-Export Principle follows immediately: According to the latter,

(4.22) $$P(\varphi \to \psi|\overline{\psi}) = P(\psi|\varphi\overline{\psi}) = 0$$

whereas according to Constancy,

(4.23) $$P(\varphi \to \psi|\overline{\psi}) = P(\psi|\varphi)P(\overline{\varphi}|\overline{\psi})$$

It is important to realize that Jeffrey has no compelling *formal* reason to require the value to be constant. To see that no harm is done by allowing it to vary across $\overline{\varphi}$-worlds, recall that all that the Thesis requires of the values at non-antecedent worlds is that their *expectation* equal $P(\psi|\varphi)$ (Fact 6, page 54). As long as that is the case, the overall values at all worlds are guaranteed to add up to $P(\psi|\varphi)$ as desired.

The equality of the expectation with the conditional probability can be satisfied by any number of ways of distributing the probability mass over the $\overline{\varphi}$-worlds. Jeffrey's (and, following him, Stalnaker's) contention that the value must in addition be *constant* is not essential and in fact leads to unwelcome consequences, as will become apparent below.

4.4.3 Interpreting compounds of conditionals

We are looking for the correct assignment of values to right-nested conditionals and conjunctions thereof. The simple rule in Definition 10 on page 57 does not cover those cases. For simplicity, I will continue to use a non-tensed language similar to $\mathcal{L}_{\mathcal{A}}$ from Definition 7, page 52, but restricted to its right-nested fragment.

Definition 34 (Language $\mathcal{L}_{\mathcal{A}}^{R}$)
The right-nested *language $\mathcal{L}_{\mathcal{A}}^{R}$ is the smallest set containing the conditional-free language $\mathcal{L}_{\mathcal{A}}^{1}$ (cf. Definition 3, p. 48) and such that for all $\varphi \in \mathcal{L}_{\mathcal{A}}^{1}$ and $\psi, \chi \in \mathcal{L}_{\mathcal{A}}^{R}$, $\varphi \to \psi$, $\overline{\psi}$, $\psi\chi \in \mathcal{L}_{\mathcal{A}}^{R}$.*

4.4.3.1 Informal characterization

Before turning to the formal definition, it is helpful to consider what the the values *should* be.

Conditional consequents. Consider first the simple case of a conditional consequent in (4.24).

(4.24) $\varphi = A \to (B \to C)$

As before, values are assigned to the worlds at which the antecedent A is true, and their expectation is subsequently "carried over" to those where A is false. At the A-worlds, the consequent $B \to C$ receives three values. It is true at ABC-worlds and false at $AB\overline{C}$-worlds.[6]

McGee and Jeffrey agree on that assignment, but disagree on the value at $A\overline{B}$-worlds. This is where the Import-Export Principle makes a difference: If it is upheld, the value should be the expectation of $AB \to C$; if not, it should be the expectation of $B \to C$. The overall assignment then should be

(4.25) a. 1 at ABC

b. 0 at $AB\overline{C}$

c. $P(C|B)$ at $A\overline{B}$ (Jeffrey)

d. $P(C|AB)$ at $A\overline{B}$ (McGee)

Finally, just like for simple conditionals, the $\overline{A}$-worlds are assigned the expectation over those worlds for which (4.25a–d) specify a value (i.e., the A-worlds,) so calculating that restricted expectation and normalizing yields the overall expectation, thus the probability of the sentence.

Conjunctions of conditionals. Consider next the conjunction in (4.26).

(4.26) $\varphi = (A \to B) \wedge (C \to D)$

There are worlds at which both conditionals are true (the $ABCD$-worlds) and ones where one or both are false ($A\overline{B}$ and $C\overline{D}$). At those worlds, the values should be 1 and 0, respectively. At others, one conjunct is true and the other has a false antecedent ($AB\overline{C}$ and $\overline{A}CD$), or both have false antecedents ($\overline{A}\,\overline{C}$). For these

[6] I argued above that the Import-Export Principle is not valid. The difference between the approaches does not concern the distribution of *truth values*: They are true or false at the same worlds. In that sense, it can be said that they roughly "mean the same." The difference lies in the distribution of the intermediate values.

cases, we may refer with McGee (1989) to the behavior of a rational betting agent and assign at worlds where one conjunct is true the probability that the other is true as well. In summary, the values at the various worlds should be

(4.27) a. 1 in $ABCD$

b. 0 in $A\overline{B} \vee C\overline{D}$

c. $P(D|C)$ in $AB\overline{C}$

d. $P(B|A)$ in $\overline{A}CD$

As in the case of conditional consequents, (4.27a–d) do not specify values for all worlds. In calculating the probability of the conjunction, however, it is sufficient to take the expectation only over those worlds for which they do, then normalize.

4.4.3.2 Formal definition

The general definition of values assigned to arbitrary sentences of the right-nested language $\mathcal{L}^R_\mathcal{A}$ involves some complications. The reason is that conjunctions which include multiple conditionals with false antecedents cannot be "taken apart." For instance, when given a conjunction $(A \to B) \wedge (C \to D)$ of two conditionals both of whose antecedents are false at the world in question, the two conjuncts cannot be treated individually, since there would be no straightforward way to "merge" the two resulting values for $A \to B$ and $C \to D$. For instance, if A and C are incompatible, multiplying the values appears to be a good choice. However, if the conjunction is of the form $(A \to C) \wedge (A \to C)$, multiplication is clearly not correct.

Thus compounds which include conditionals with false antecedents must not be decomposed; the expectation must be calculated for an expression holding both. On the other hand, non-conditional constituents and conditionals with true antecedents are evaluated directly at the world in question.

In Definition 37 below, I formalize a procedure suggested by Stalnaker and Jeffrey (1994): Sentences in $\mathcal{L}^R_\mathcal{A}$ are first replaced with simpler sentences in $\mathcal{L}^R_\mathcal{A}$, evaluating non-conditional constituents and replacing conditionals whose antecedents are true at the world in question with their consequents. (Since $\mathcal{L}^R_\mathcal{A}$ allows only right-nested conditionals, antecedents are guaranteed to have truth values everywhere).[7]

Definition 37 improves over Stalnaker and Jeffrey (1994) in that the values assigned to compounds including conditionals with false antecedents are actually constructed. Stalnaker and Jeffrey (1994) instead impose a consistency condition on probability distributions to the effect that those values must equal the overall expectation—which cannot be calculated without those very values. They state that that condition "makes no assumption about what the value is in this case—just that whatever it is, it is the same as the overall expectation of that random variable" (p. 37) and correctly point out that whenever there is a value which satisfies the constraint, it is unique.

The evaluation is most easily defined by introducing two auxiliary devices. The first is a pair of symbols abbreviating arbitrary tautologies and contradictions.

Definition 35 (Notational convention)
Let the symbols $\top$ and $\bot$ abbreviate arbitrary tautologies and contradictions in the conditional-free language $\mathcal{L}^1_\mathcal{A}$. I.e., for all interpretations V^1 and all worlds w: $V(\top)(w) = 1$ and $V(\bot)(w) = 0$.

I use the two symbols merely for perspicuity. One may instead replace them with actual sentences in the language, such as $A \vee \neg A$ and $A \wedge \neg A$, respectively.

The second auxiliary device is a world-dependent syntactic transformation, hinted at above, which simplifies expressions in the right-nested language $\mathcal{L}^R_\mathcal{A}$. The goal is to reduce them as far as possible, evaluating all constituents that need to be evaluated locally, i.e., all constituents not dominated by a conditional with a false antecedent.

Definition 36 (Simplification map)
Given a truth assignment V^1 of sentences in the non-conditional language $\mathcal{L}^1_\mathcal{A}$, a family of functions π from $\mathcal{L}^R_\mathcal{A}$ into $\mathcal{L}^R_\mathcal{A}$ is defined as follows: For each world $w \in W$

[7] It is possible to reach the same values without this syntactic transformation. I will outline that option in the Appendix. There is a tradeoff, however: The syntactic ease is bought at the cost of a considerable complication of the model-theoretic apparatus. For now, I will avoid this complexity.

and sentence $\varphi \in \mathcal{L}^R_{\mathcal{A}}$,

$$\textit{if } \varphi \in \mathcal{L}^1_{\mathcal{A}}\textit{: } \pi_w(\varphi) = \begin{cases} \top \textit{ if } V^1(\varphi)(w) = 1 \\ \bot \textit{ if } V^1(\varphi)(w) = 0 \end{cases}$$

$$\pi_w(\overline{\varphi}) = \begin{cases} \top \textit{ if } \pi_w(\varphi) = \bot \\ \bot \textit{ if } \pi_w(\varphi) = \top \\ \neg\varphi \textit{ otherwise} \end{cases}$$

$$\pi_w(\varphi\psi) = \begin{cases} \bot \textit{ if } \pi_w(\varphi) = \bot \textit{ or } \pi_w(\psi) = \bot \\ \varphi \textit{ if } \pi_w(\psi) = \top \\ \psi \textit{ if } \pi_w(\varphi) = \top \\ \varphi \wedge \psi \textit{ otherwise} \end{cases}$$

$$\pi_w(\varphi \to \psi) = \begin{cases} \pi_w(\psi) \textit{ if } \pi_w(\varphi) = \top \\ \varphi \to \psi \textit{ if } \pi_w(\varphi) = \bot \end{cases}$$

Notice that in the clause for conditionals, $\pi_w(\varphi)$ is guaranteed to be either $\top$ or $\bot$ since φ, being an antecedent, does not contain conditionals.

Finally, the interpretation V^R for $\mathcal{L}^R_{\mathcal{A}}$ evaluates the simplified versions of sentences locally, where that is appropriate. The output of π_w is either $\top$ or $\bot$ if the original sentence contains no conditionals with false antecedents, or if all such conditionals are conjoined with sentences false at w. Otherwise, the remaining conditionals are "sent off" for evaluation at those worlds where further simplifications are possible.

Definition 37 (Interpretation for $\mathcal{L}^R_{\mathcal{A}}$)
An interpretation $V^R : \mathcal{L}^R_{\mathcal{A}} \mapsto \mathbb{R}^W$ *is a function from sentences in* $\mathcal{L}^R_{\mathcal{A}}$ *to random variables, satisfying the following conditions for all sentences* φ *in* $\mathcal{L}^R_{\mathcal{A}}$ *at all* $w \in W$*:*

$$V^R(\varphi)(w) = \begin{cases} 1 \textit{ if } \pi_w(\varphi) = \top \\ 0 \textit{ if } \pi_w(\varphi) = \bot \\ E[V^R(\pi_w(\varphi))|\pi(\pi_w(\varphi)) \neq \pi_w(\varphi)] \textit{ otherwise} \end{cases}$$

Definition 37 defines values for most sentences, *except* for ones containing conditionals whose antecedents are false (almost) everywhere. In such a case, the conditional expectation is taken over a set of zero (or infinitesimal) probability, so that any arbitrary value can be assigned without affecting the overall expectation. As I suggested earlier, in such cases the value should be taken to be undefined, rendering the use of the sentence infelicitous.[8]

In terms of this syntactic transformation, the difference between the Jeffrey-style conditional defined above and McGee's comes in at the point where π_w replaces conditionals with true antecedents. According to the Import-Export Principle, at a world where the antecedent is true, all further expectations taken in the course of the interpretation of the consequent, if any, must be restricted to antecedent-worlds. To enforce this in the present setup, a somewhat more complex mapping π' would be needed according to which

$$\pi'_w(\varphi \to \psi) = \pi'_w(\chi) \tag{4.28}$$

if $\pi_w(\varphi) = \top$, where χ is the result of conjoining φ with the antecedents of all conditionals embedded in ψ.

4.4.3.3 Illustrations

To see the definitions in the preceding section at work, I will step through the interpretation in some detail using simple examples.

Simple conditionals. First, consider a simple sentence of the form

(4.29) $\varphi = A \to C$

[8] Another popular way to take care of this case is by stipulation. Adams (1965), McGee (1989) and others set the value to 1, with the result that all probabilistically valid inferences are also classically valid.

both of whose constituents are conditional-free. At a world w at which A is true, π_w reduces φ to C and the truth value of C at w is assigned as the value of φ. At a world w' where A is false, the value is the expectation of its unreduced version $A \to C$ over all the worlds where it can be reduced, that is, the A-worlds. At those worlds it is again reduced to C, so the values are

(4.30) a. 1 in AC

b. 0 in $A\overline{C}$

c. $E[V^R(A \to C)|A] = P(C|A)$ in $\overline{A}$

It is worthwhile to spell out the calculation for the overall expectation in some detail, even though doing so is a bit tedious.

Proposition 38 (Simple conditionals)
If $\varphi = A \to C$, where A, C are in $\mathcal{L}^1_{\mathcal{A}}$, then $E[V^R(\varphi)] = P(C|A)$[9]

Conditional consequents. Next, the sentence

(4.31) $\varphi = A \to (B \to C)$

is interpreted as follows: At a world w where A is true, it is reduced by π_w to $B \to C$. If B is true at w as well, it is further reduced to C and receives the truth value C has at w. If, instead, B is false at w, the expectation of $B \to C$ over all those worlds where further simplifications are possible (i.e., the B-worlds) is taken. At those worlds, the value of $B \to C$ is that of C, so that the value assigned at w is $P(C|B)$.

At worlds where A is false, on the other hand, the expectation of the unaltered conditional φ over those worlds where it can be simplified (i.e., the A-worlds) is taken. Thus the values are

[9] *Proof.*
$$\begin{aligned} E[V^R(\varphi)] &= E[V^R(\varphi)|A]P(A) + E[V^R(\varphi)|\overline{A}]P(\overline{A}) \\ &= E[V^R(C)|A]\big(P(A) + P(\overline{A})\big) = P(C|A) \end{aligned}$$
□

(4.32) a. 1 in ABC

b. 0 in $AB\overline{C}$

c. $E[V^R(\varphi)|A]$ in $\overline{A}$

The expectation of those values is given in Proposition 39.

Proposition 39 (Conditional consequents)
If $\varphi = A \to (B \to C)$, where $A, B, C \in \mathcal{L}^1_{\mathcal{A}}$, then[10]

$$E[V^R(\varphi)] = P(BC|A) + P(C|B)P(\overline{B}|A)$$

Conjunctions of conditionals. Finally, consider sentence in (4.33), again made up of non-conditional components.

(4.33) $\varphi = (A \to B) \wedge (C \to D)$

If (4.33) is interpreted at a world w where both A and C are true, it is reduced by π_w to $B \wedge D$, which, being conditional-free, are evaluated by V^1 and replaced with their respective truth values. The result is that the value of $V^R(\varphi)$ is

(4.34) a. 1 in $ABCD$

b. 0 in $A\overline{B}C \vee AC\overline{D}$

At a world w' where A is true and C is false, the value of B determines whether any additional steps are necessary. $\pi_{w'}$ reduces φ first to $B \wedge (C \to D)$, then to

[10] *Proof.*
$$\begin{aligned}
E[V^R(\varphi)] &= E[V^R(\varphi)|A]P(A) + E[V^R(\varphi)|A]P(\overline{A})\\
&= E[V^R(\varphi)|A]\\
&= \textstyle\sum_x x \cdot P(\{w|V^R(\varphi)(w) = x\}|A)\\
&= 1/P(A) \cdot \textstyle\sum_x x \cdot Pr(\{w|V^R(\varphi)(w) = x\} \cap A)\\
&= 1/P(A) \cdot [0 \cdot P(AB\overline{C}) + 1 \cdot P(ABC) + E[V^R(B \to C)]P(A\overline{B})]\\
&= 1/P(A) \cdot \big(P(ABC) + P(C|B)P(A\overline{B})\big)\\
&= P(BC|A) + P(C|B)P(\overline{B}|A)
\end{aligned}$$
□

either $C \to D$ or $\bot$, depending on the value of B at w'. If the value is $C \to D$, its expectation over those worlds is taken in which C is true (i.e., where a further simplification of $C \to D$ to D succeeds. Likewise for worlds where A is false and C is true. So the values of φ are

(4.35) a. 0 in $A\overline{B}\,\overline{C} \vee \overline{A}C\overline{D}$

b. $E[V^R(C \to D)|C] = P(D|C)$ in $AB\overline{C}$

c. $E[V^R(A \to B)|A] = P(B|A)$ in $\overline{A}CD$

Finally, where both A and C are false, φ is left unmodified and its conditional expectation over all other worlds is taken:

(4.36) $E[V^R(\varphi)|A \vee C]$ in $\overline{A}\,\overline{C}$

Given the other clauses, that value is well-defined (unless A or C has zero probability). The unconditional overall expectation of φ can again be spelled out in the tedious format.

Proposition 40 (Conjoined conditionals)
If $\varphi = (A \to B) \wedge (C \to D)$, where $A, B, C, D \in \mathcal{L}^1_{\mathcal{A}}$, then[11]

$$E[V^R(\varphi)] = \frac{P(ABCD) + P(D|C)P(AB\overline{C}) + P(B|A)P(AB\overline{C})}{P(A \vee C)}$$

[11] *Proof.*
$$\begin{aligned}
E[V^R(\varphi)] &= E[V^R(\varphi)|A \vee C]P(A \vee C) + E[V^R(\varphi)|A \vee C]P(\overline{A}\,\overline{C}) \\
&= E[V^R(\varphi)|A \vee C] \\
&= 1/P(A \vee C) \cdot \textstyle\sum_x x \cdot Pr(\{w|V^R(\varphi)(w) = x\} \cap A \vee C) \\
&= 1/P(A \vee C) \cdot \begin{bmatrix} 0 \cdot (P(A\overline{B}) + P(C\overline{D}))+ \\ 1 \cdot P(ABCD)+ \\ E[V^R(A \to B)|A]P(\overline{A}CD)+ \\ E[V^R(C \to D)|C]P(AB\overline{C}) \end{bmatrix} \\
&= 1/P(A \vee C) \cdot \big(P(ABCD) + P(B|A)P(\overline{A}CD) + P(D|C)P(AB\overline{C})\big)
\end{aligned}$$
□

4.4.4 The values at non-antecedent worlds

With these definitions in place, we can return to the question of what the values of conditionals at non-antecedent worlds should be. The Import-Export Principle requires them to depend on the value of the consequent at those worlds. Constancy, on the other hand, requires them to be uniformly distributed.

In this section I will address this question by considering examples that have been brought forth in the literature to argue that the definitions make counterintuitive predictions. The first suggests that right-nested conditionals ought to observe the Import-Export Principle, *contra* Jeffrey. The second suggests that the values of conjunctions of conditionals, on which both McGee and Jeffrey agree, are wrong. The Import-Export Principle is irrelevant to this case; however, intuition suggests a solution. That solution, it turns out, also solves the first problem because it has a similar effect as the Import-Export Principle. The formal implementation of the solution will have to wait until the next chapter.

4.4.4.1 Conditional consequents

Under Constancy, the value assigned to a conditional with conditional consequent is as in (4.37).

(4.37) $$P(A \to (B \to C) = P(BC|A) + P(C|B)P(\overline{B}|A)$$

(Edgington, 1991, p. 200) challenges this prediction with the following example. A match is about to be struck. It may be wet with probability .55, and it certainly won't light if it is wet. The probability that it will be struck is .5. The probability that it will light if struck is .9. The match is wet if it is struck and does not light, or if it is not struck. The numbers are as follows:

(4.38) S The match is struck $P(S)$=.5
W The match is wet $P(W)$=.55
$P(W|S\overline{L})=P(W|\overline{S}) = 1$
L The match lights $P(L|W)$=0
$P(L|S)$=.9

The sentence in question is given in (4.39a), its formal representation in (4.39b).

(4.39) a. If it is wet, then if you strike it, it will light.
b. $W \rightarrow (S \rightarrow L)$

Intuitively this sentence is false and should receive zero probability. That is not, however, what Jeffrey's formula predicts. To apply the equation in (4.37), we only need to calculate

$$P(\overline{S}|W) = \frac{P(W|\overline{S})P(\overline{S})}{P(W)} = \frac{1 \cdot .5}{.55} \approx .91 \quad (4.40)$$

Since $P(L|W)$ is zero, so is $P(LS|W)$. The resulting value is far too high:

$$\begin{aligned} P(W \rightarrow (S \rightarrow L)) =& P(LS|W) + P(L|S)P(\overline{S}|W) \\ =& 0 + .9 \cdot .91 \\ \approx& .82 \end{aligned} \quad (4.41)$$

The first comment to be made about Edgington's figures is the following: The scenario is set up in such a way that the match is wet if it is not struck. In other words, whether the match is struck or not depends on whether it is wet. Note that this assumption is not needed to make the point: Without such a dependency, the value for $P(W|\overline{S})$ in (4.40) would be less than 1, but the result of substituting it in (4.41) would still be well above 0.

However, the example does support Edgington's point. It is true that the McGee conditional would yield the desired result:

(4.42) $$P(W \to (S \to L)) = P(L|SW) = 0$$

What went wrong? Intuitively, the problem is this: The non-zero term in the right-hand side of (4.41) is the conditional expectation, restricted to W, of the values assigned at those worlds where S is false. The value assigned is $P(L|S)$; its weight in the overall sum is $P(\overline{S}|W)$. In words, it is the value of (4.43) at worlds where the match is both wet and not struck.

(4.43) If the match is struck, it will light

But this is not intuitively correct. The match is wet, and it is not struck. What is the probability that it *would* light if it *were* struck? Clearly not the unconditional probability that it would light if struck. The fact that it is wet would remain, even if it had been struck, and it is relevant to the value of the conditional. The value should be $P(L|SW)$, rather than $P(L|S)$. However, following Constancy, the overall probability that the match lights if struck, $P(L|S) = .9$, is evenly spread over *all* worlds in which it is not struck, regardless of whether it is wet in those worlds or not.

The problem would not arise if the random variable $V(S \to L)$ were made sensitive to such circumstantial factors, which are not explicitly mentioned in the antecedent of the conditional but would remain unchanged in those worlds where its antecedent is true. At the $\overline{S}$-worlds, the value should not be uniformly distributed, but vary depending on whether those worlds are in $\overline{S}$W or in $\overline{S}\,\overline{W}$.

This intuition needs refining. Other sentences are true at those worlds as well. For instance, it is a fact at all worlds in $W\overline{S}$ that the match does not light. Why not take *that* into consideration and assign instead $P(L|S\overline{L})$? The effect would be the same, as $P(L|S\overline{L})$, too, equals 0.

This latter suggestion seems absurd. It is clear that at worlds where the match is wet and not struck, (4.43) is false, not because the match does not light, but because it is wet.

The word *because* in the last sentence gives a hint at the solution. Whether the match lights if struck depends on whether it is wet or dry, not whether it in fact lights or not. The dependence is *causal*. It enters the picture as a general assumption about relationships between facts. Despite the fact that its effect is similar to that of the Import-Export Principle, it has nothing to do with the antecedent of the conditional. For suppose it did arise from the Import-Export Principle and the McGee conditional in (4.42) were indeed correct. Then the sentence in (4.44a), paraphrased for clarity in (4.44b), should be false:

(4.44) a. If it does not light, then if you strike it, it will light.

b. If it is true that it does not light, then it is also true that if you strike it, it will light.

c. $P(L|S\overline{L}) = 0$

In Edgington's scenario, (4.44a) is indeed false, but that is solely due to the assumption that the match is wet if it is not struck. Without that eccentric stipulation, (4.44a) does not appear all that false anymore: Most worlds in which the match does not light are ones in which it is not struck. The consequent of (4.44a) is not false at those worlds.

4.4.4.2 Conjunctions of conditionals

Another reason to locate the source of the relevance of the wetness of the match in facts about causal relations, rather than in the sentence itself, is that the same assumption helps explain why the values assigned by both Jeffrey and McGee to conjunctions of conditionals are sometimes counterintuitive.

Lance (1991) discusses the following example: A werewolf lives nearby who under full moon, such as tonight, always stalks one of two neighborhoods, n or m, completely at random. Anybody who goes outside in the neighborhood stalked by the werewolf is killed with absolute certainty. We are in a house in neighborhood n. Suddenly we realize that Jones, who was in the house earlier, hasn't been seen for a while. There is a fifty percent chance that he left the house, but perhaps he is upstairs. There are two doors through which he might have left, a and c. We have no reason to assume that Jones died for any reason but the werewolf. To summarize, the following are the probabilities as described in the scenario:

(4.45)	N	"The werewolf is stalking n"	$P(N)$=.5
	M	"The werewolf is stalking m"	$P(M)$=.5
	S	"Jones is inside"	$P(S)$=.5
	A	"Jones left through door a"	$P(A)$=.25
	C	"Jones left through door c"	$P(C)$=.25
	K	"Jones is killed"	$P(K \mid (A \vee C) \wedge N)$=1

Now consider the sentence in (4.46a), translated into the expression in (4.46b).

(4.46) a. If Jones left through door a he was killed, and if Jones left through door c he was killed.

b. $\varphi = (A \rightarrow K) \wedge (C \rightarrow K)$

Intuitively, φ is certainly true in those worlds where the werewolf is stalking n and certainly false where the werewolf is stalking m. So its truth fully depends on whether the werewolf is stalking our neighborhood, that is, intuitively one would expect that $P(\varphi) = .5$. Now $P(AKC) = 0$, since Jones cannot have left through both doors. By Constancy, the intermediate value to be assigned to $A \rightarrow K$ at those worlds where Jones did not leave through door a is $P(K|A)$. Likewise, $P(K|C)$ is assigned to $C \rightarrow K$ at the worlds where he did not leave through door c. In the scenario, $P(K|A) = P(K|C) = .5$. The probability that he left through door c and was killed is $P(CK) = P(\overline{A}KC) = .125$; similarly, $P(AK) = P(AK\overline{C}) = .125$. Now according to the formula,

(4.47)
$$P(\varphi) = \frac{P(AKC) + P(A \to K)P(\overline{A}KC) + P(C \to K)P(AK\overline{C})}{P(A \vee C)}$$
$$= \frac{0 + .5 \cdot .125 + .5 \cdot .125}{.5} = .25$$

This is only half of the intuitively correct .5. What went wrong? Lance made some suggestive remarks in his discussion, without however reaching a conclusion or proposing a repair. Put briefly, the problem is this: Consider the worlds in $\overline{A}KC$, where Jones left through door C and was killed. The value assigned to (4.48a) at those worlds is $P(A \to K) = P(K|A)$.

(4.48) a. If Jones left through door a, he was killed.
b. $A \to K$

What are the facts at those worlds? Jones leaves through door c and is killed; that can happen only if the werewolf is stalking neighborhood n. So all worlds in $\overline{A}KC$ are N-worlds: $P(N|KC) = 1$. Most crucially, Jones' behavior does not affect the werewolf's behavior: Whether Jones leaves through door a or c, or stays inside, the werewolf still stalks whichever neighborhood she stalks. Notice that $P(K|AN)$, the probability that Jones is killed, given that he left through door a and the werewolf is stalking neighborhood n, is 1. That should, intuitively, be the value (4.48a) has at those worlds. However, the formula assigns $P(K|A) = .5$, the expectation taken over *all* worlds, those in N and those in M.

Likewise, at the worlds where Jones leaves through door c and is *not* killed, the werewolf must be stalking neighborhood m. But then, if Jones had left through door a instead, he would not have been killed either. The value of (4.48a) at those worlds should be $P(K|AM) = 0$, rather than $P(K|A) = .5$.

The situation is mirrored at those worlds where Jones leaves through door a. Where he is killed, he would certainly have been killed if he had left through door c instead. Where he is not killed, he would not have been killed if he had left through door c.

This discussion again suggests a systematic way in which the problem can be circumvented. Suppose $V(A \to K)$ "recognizes" the fact that an $\overline{A}N$-world would still be an N-world, even if Jones had left through door a. It would "harvest" the value assigned at the $\overline{A}N$-worlds to the conditional $A \to K$ by taking the expectation only over the AN-worlds, not all A-worlds, hence the value would be $P(K|AN)$, rather than $P(K|A)$. Similarly for the C-worlds. Then

(4.49)
$$\begin{aligned} P(\varphi) &= \frac{P(AKC) + P(K|AN)P(\overline{A}KC) + P(K|CN)P(AK\overline{C})}{.5} \\ &= \frac{0 + 1 \cdot P(\overline{A}KC) + 1 \cdot P(AK\overline{C})}{.5} \\ &= \frac{.125 + .125}{.5} = .5 \end{aligned}$$

We can expand on Lance's example and consider the following sentence:

(4.50) a. If Jones left through door a he was killed, and if he left through door c he was *not* killed.

b. $\psi = (A \to K) \wedge (C \to \overline{K})$

To the extent that it is possible to have intuitions about (4.50), I strongly feel that it is false, hence its probability ought to be 0. However, the formula predicts otherwise:

(4.51)
$$\begin{aligned} P(\psi) &= \frac{P(AKC\overline{K}) + P(K|A)P(\overline{A}C\overline{K}) + P(\overline{K}|C)P(AK\overline{C})}{.5} \\ &= \frac{0 + .5 \cdot .125 + .5 \cdot .125}{.5} = .25 \end{aligned}$$

The nature of the problem is the same as before: At worlds where Jones leaves through door c and is not killed, he would not have been killed if he had left through door a, for those worlds are M-worlds and would remain so if Jones had left through door a. Intuitively, the value assigned to (4.52a) at such a worlds should be 0.

(4.52) a. If Jones left through door a, he was killed.

b. $A \to K$

That is precisely what the formula would predict if it were "sensitive," in the above sense, to the relevance of M and N to the value of the conditional:

(4.53)
$$\begin{aligned} P(\psi) &= \frac{P(ACK\overline{K}) + P(K|AM)P(\overline{A}C\overline{K}) + P(\overline{K}|CN)P(AK\overline{C})}{.5} \\ &= \frac{0 + 0 \cdot .125 + 0 \cdot .125}{.5} = 0 \end{aligned}$$

4.5 Summary

The upshot of the discussion so far is by now clear: The value at non-antecedent worlds should not be constant, but allowed to vary depending on factors that are not affected by the truth or falsehood of the antecedent. Thus Constancy is not tenable. Secondly, what those factors are cannot be determined from the sentence itself via the Import-Export Principle. Not only would that lead to the wrong predictions for right-nested conditionals, but the need to take such factors into accounts arises also in sentences which do not even mention the needed additional information, such as conjunctions of conditionals.

Thirdly, in the discussion I have appealed a number of times to intuitions about counterfactuals: At the N-worlds, Jones *would have* been eaten if he *had left* the house, at the M-worlds he *would* not. There clearly is a close connection between the two. Causal relations of the kind I propose here to supplant the Import-Export Principle have been recognized to play a vital part in the interpretation of counterfactuals. We have seen here that their role is just as important in predictive conditionals.

Chapter 5

Counterfactuals and causality

5.1 Introduction **119**
5.2 Counterfactuals **119**
5.2.1 Past predominance 119
5.2.2 Overall similarity 121
5.2.3 Posterior facts 122
5.2.4 Causal independence 127
5.3 Causality **133**
5.3.1 What causal relations relate 134
5.3.2 Particular and general causation 135
5.3.3 Deterministic causation of probabilities 136
5.3.4 Implementation 137
5.4 Applications **140**
5.4.1 Right-nested conditionals 140
5.4.2 Conjunctions of conditionals 145
5.4.3 Counterfactuals: Prior propensity 146
5.4.4 Import-Export revisited 151
5.5 Summary **152**

5.1 Introduction

In the preceding chapter I repeatedly alluded to the relevance of intuitions about *counterfactual* conditionals and *causality* to the interpretation of embedded and compounded conditionals. It is time to take a closer look at both of these in their own right.

In this chapter I am going to survey the relevant data about counterfactuals, motivate a formal implementation of causal relations in the models, and spell out in some more detail the intuition that present counterfactuals are equivalent to past predictive conditionals. The intuition itself has been appealed to in many writings on counterfactuals (Adams, 1975; Skyrms, 1981; Thomason and Gupta, 1981; van Fraasen, 1981; Edgington, 1995; Mårtensson, 1999, and others), although in the form it is usually stated it is not be quite correct (cf. Barker, 1998).

5.2 Counterfactuals

Counterfactuals raise many questions. This chapter is in no way intended to address them all. I only touch on a few relevant issues.

The basic motivating intuition behind most accounts of counterfactuals, whether phrased in probabilistic terms or not, is this: The interpretation of conditionals whose antecedent is false must be based upon some operation of making worlds "accessible" in which it is true, but restricting the accessibility to those that are "similar" in some sense.

5.2.1 Past predominance

In dealing with counterfactuals in time, the right kind of "similarity" is often defined as some version of the principle of *Past Predominance* (Thomason and Gupta, 1981):[1]

[1]Lest I misrepresent the ideas of Thomason and Gupta, I should point out that later in the same article they do impose a restriction of the choice of possible futures to ones which are "causally

(5.1) In determining how close [w at time t] is to [w' at time t], past closeness predominates over future closeness; that is, the portions of w and w' not after t predominate over the rest of w and w'. (p. 301)

Thus those worlds are "closest" to a given world which share the longest initial segment of history with the latter. So the way to find the closest antecedent-worlds is to "unravel" history up to the latest earlier time at which it was still possible for the antecedent to be true, then examine, from the perspective of that point, the continuations in which it in fact *was* true. A well-known definition in this spirit is due to Lewis (1979):[2]

Definition 41 (Past predominance (Lewis, 1979))
Consider a counterfactual 'If it were that A, then it would be that C' where A is entirely about affairs in a stretch of time t_A. Consider all those possible worlds w such that:

a. *A is true at w;*

b. *w is exactly like our actual world at all times before a transition period beginning shortly before t_A;*

c. *w conforms to the actual laws of nature at all times after t_A; and*

d. *during t_A and the preceding transition periods, w differs no more from our actual world than it must to permit A to hold.*

The counterfactual is true if and only if C holds at every such world w.

There are obvious parallels between Definition 41 and the assignment of probabilistic values of simple predictive conditionals defined in Chapter 4: We are assigning, at time t, the value of a conditional '*if* A *then* C', where A is about a *later* time t',

coherent" in order to address problems of the kind discussed below.

[2]Lewis in fact does not endorse this definition as generally applicable. I am not going to discuss his reasons in detail here; however, Lewis does accept Definition 41 as basically right where applicable, and consequently has to add a number of problematic stipulations in order to circumvent the kind of problems discussed here.

at a world w where A is false at t'. The world w is a member of the set $h(t)$ of worlds which agree on all facts up to time t. The set of A-worlds to be considered consists of those which "grow out of" $h(t)$. The last clause may be replaced with (5.2):

(5.2) The value of the conditional is the conditional expectation of the value of C over the set of A-worlds in $h(t)$.

That value is 1 whenever the universally quantified condition holds, but can in general be short of 1.[3]

The problem with this account, both for counterfactuals and for predictive conditionals, and whether it be couched in probabilistic terms or not, is that it "undoes" the outcomes of processes which become settled independently of the value of the antecedent during the "unraveled" segment of history. In the following, I will briefly discuss a few examples to illustrate the consequences.

5.2.2 Overall similarity

I mentioned in Footnote 2 on page 120 that Lewis (1979) did not consider the analysis in Definition 41 sufficiently general. It does not capture, he argued, counterfactuals which do not in any obvious way depend for their interpretation on the temporal order in which the facts came about. To extend the definition to the non-temporal case, as well as to take care of some more exotic notions of time, he instead proposed a more general definition:

Definition 42 (Overall similarity (Lewis, 1979))
A counterfactual 'If it were that A, then it would be that C' is (non-vacuously) true if and only if some (accessible) world where both A and C are true is more similar to our actual world, overall, than is any world where A is true but C is false.

[3] The converse does not hold: The conditional probability may be 1 while the universally quantified condition does *not* hold, in case there is a non-empty set of $A\overline{C}$-worlds with zero probability; but that complication need not concern us here.

Lewis intended Definition 42 as a general scaffolding which can be fleshed out in various ways, one of which is Definition 41, where the vague notion of "overall similarity" is given some more content by reference to temporal precedence.

It has been pointed out, however, that the two definitions are in fact at odds with each other and that neither one fully accounts for a variety of problematic cases. The problem is that past predominance is *not* a special case of "overall similarity" in any straightforward and intuitive sense. Past predominance measures similarity as similarity over an initial sub-history. In contrast, overall similarity is just that—*overall* similarity.

5.2.3 Posterior facts

Thus a difference arises with respect to facts at times *later* than that referred to in the antecedent. In the absence of further precisifications, past predominance is blind to those posterior facts, while overall similarity gives them as much weight as past facts. Neither one is correct: Posterior facts sometimes do and sometimes do not affect the value of the counterfactual, as the following examples demonstrate.

5.2.3.1 Irrelevant posterior facts

Fine (1975) presented a well-known counterexample to the overall similarity account, interpreted as measuring similarity among *all* facts throughout all of history:

> The counterfactual [(5.3)] is true or can be imagined to be so.
>
> (5.3) If Nixon had pressed the button there would have been a nuclear holocaust
>
> Now suppose that there never will be a nuclear holocaust. Then that counterfactual, on Lewis' analysis, very likely false. For given any world in which antecedent and consequent are both true it will be easy to imagine a closer world in which the antecedent is true and the consequent false.

> For we need only imagine a change that prevents the holocaust but that does not require such a great divergence from reality. (p. 452)

Fine's example speaks in favor of past predominance and against overall similarity. Small changes may cause vast divergences between worlds, but what counts is the history up to the small change, which together with the change is all that is needed to evaluate the counterfactual.

Lewis (1979) goes at some length to suggest how examples like these could be reconciled with his preference for overall similarity, introducing a hierarchy among the various "miracles" that would lead from the world of interpretation to one where the antecedent is true:

(5.4)
a. It is of first importance to avoid big, widespread, diverse violations of law.
b. It is of the second importance to maximize the spatio-temporal region throughout which perfect match of particular fact prevails.
c. It is of the third importance to avoid even small, localized, simple violations of law.
d. It is of little or no importance to secure approximate similarity of particular fact, even in matters that concern us greatly.

One can view these rules as a re-introduction of a "dampened" version of past predominance, which, in a sense, is exactly what is needed. The "small" *divergence* miracle which would cause Nixon to press the button is more admissible than the "large" *convergence* miracle that would be required to restore similarity in particular fact after the button is pushed, since all traces and consequences of Nixon's hypothetical act would have to be obliterated.[4] Perfect match in particular fact could therefore only be attained during that segment of history which precedes his (actual) non-pushing of the button. Later differences in particular fact, though vast between the two courses of history, count little towards their dissimilarity as they are outweighed by obedience of the laws.

[4] The terms "divergence miracle" and "convergence miracle" are due to Nute (1980).

Many authors have complained that Lewis' notion of "miracle" is too vague to serve as a useful guideline and makes it difficult to test the predictions of the theory (Nute, 1980, 1984, and elsewhere). I share this sentiment and will present a somewhat more explicit version below.

5.2.3.2 Relevant posterior facts

The preceding section suggested that facts which hold later than the time referred to in the antecedent of a counterfactual should not count towards similarity, thus should not be "held constant" in collecting the relevant set of antecedent-worlds. That is not quite correct, however: Some posterior facts must be held constant as well, as the examples in this section show.

Airplane crashes. Dorothy Edgington (p.c.) has the following example: I was on my way to the airport this morning when my car broke down. Now I am back home, upset because

(5.5) If I had not missed my plane, I would be on my way to Vancouver now.

A past predominance interpretation procedure for this example involves the consideration of a time at which it was still possible that I would catch my plane, i.e., before the car broke down, and calculating the likelihood, at the time, that I would be on my way to Vancouver, given that I catch the plane. (5.5) is correctly predicted to be true.

Now, however, it turns out that in fact the plane crashed and everyone on board was killed. So I was lucky, for

(5.6) If I had not missed my plane, I would be dead now.

The interpretation procedure does not yield the correct prediction in this case: The earlier time at which it was possible that I would catch the plane is the same as before, and at that time it was extremely unlikely that the plane would crash. So I would most likely be well on my way and (5.6) is wrongly predicted to be true.

Thus although the crash occurred later than the time at which I missed the plane, it should be "held constant" in examining alternative courses of events. That is not generally the case, however: Suppose the crash was due to a miscommunication in the cockpit, which could have been avoided if a semanticist had been present. In that case, not only would I be on my way to Vancouver, but a linguist would, for once, have saved lives. My presence on the plane would have interfered with the process that brought about the crash, and (5.5) is true.

Coin tosses. Slote (1978) attributes the following example to Sidney Morgenbesser. I cite it here in the version of Bennett (1984):

(5.7) At T_1 I bet that when the coin is tossed at T_2 it will come up heads; and in the upshot it does just that; but this is a purely chance event, with no causally sufficient prior conditions. Now consider the conditional

(5.8) If I had bet on tails at T_1 I would have lost.

I agree with Bennett's judgment and that of his informants that (5.8) should come out true. However, that is not what the "unraveling" account predicts. The problem is the same as before: In worlds which "grow out of" the same history as the actual one at T_1, the chance process of tossing the coin is not yet performed. The prior chance of tails is not zero, not even small, hence the counterfactual is wrongly predicted to be false.

Tichý's Puzzle. The following is a variation on an example due to Tichy (1976) and frequently discussed in other places (Veltman, 1985, and others). It is a useful example which, when viewed from various different angles, summarizes all the points I have made so far.

(5.9) Consider a man—call him Jones—who is possessed of the following disposition as regards wearing his hat. Each morning he flips a coin before he opens the curtains to see what the weather is like. 'Heads' means he is going to wear his hat if the weather is fine, 'tails' means he is not going to wear his hat in that case. Bad weather, however, invariably induces him to wear a hat.

This morning, 'heads' came up when he flipped the coin; furthermore, the weather is bad, so Jones is wearing his hat; and

(5.10) If the weather were fine, Jones would be wearing his hat.

Had the weather been fine, the coin would nevertheless have come up heads. Thus (5.10) is correctly predicted true under past predominance since Jones tossed the coin before he knew what the weather was like.

Consider now the following minor variant of the story:

(5.11) This morning, Jones was too impatient to toss the coin first. He opened the curtains and saw that the weather was bad. But he tossed the coin all the same, since doing so is a time-honored habit of his. The coin came up 'heads'. Later in the day, he is wearing his hat.

Clearly in this case, (5.10) should still be true. However, this is no longer predicted by Definition 41: When Jones found out about the weather, he had not yet tossed the coin. A re-run of history from that point onwards would involve a different toss of the coin whose outcome could well turn out different.

The conditional would *correctly* predicted by the definition to be false if the coin toss itself depended on the weather, as in the following variant:

(5.12) Jones always opens the curtains first. If the weather is bad, he will wear his hat. If it is good, he tosses the coin. This morning the weather was bad, so he is wearing his hat.

What if the *consequences* of the outcome depend on the weather? Consider the following variation on Tichý's theme:

(5.13) Jones always opens the curtains and then tosses the coin. In bad weather, the coin decides which of his two bad-weather hats he will wear—'heads' for blue, 'tails' for black. In good weather, the toss decides whether he will wear his straw hat ('heads') or none at all ('tails'). Today the weather is bad and the coin came up 'heads', so he is wearing his blue hat.

It is clear that (5.10) is true in this scenario as well. If the weather were fine, the outcome would still have been 'heads' and it would, jointly with the weather, have induced Jones to wear his straw hat.

Notice, however, that this intuition cannot be ascribed to the fact that the *outcome* of the toss does not depend on the weather: What counts is the independence of the toss itself. For suppose we are told in addition that in bad weather Jones tosses the coin in the living room, whereas in good weather he tosses it on the balcony. Now it is no longer the case that in good weather it would have been the *very same* toss which would have decided, so the possibility that it might have come up 'tails' is again open, and (5.10) is false.

5.2.4 Causal independence

In view of examples such as the ones in the preceding section, theories of counterfactuals typically incorporate *causal* information to determine which of the posterior facts to take into account.[5]

5.2.4.1 Causal trails

The way causal information ought to be exploited is intuitively as follows: As before, history is "unraveled" up to a point at which the (now false) antecedent had non-zero

[5]Some terms and quotations in this section are from Mårtensson (1999) because his is a contemporary theory which amalgamates and builds on a large amount of earlier work.

probability. However, history is not simply re-run from there, for not *all* antecedent-worlds possible at the earlier time are in fact alternatives to the actual course of events. Instead, further restrictions are imposed upon the set of antecedent-worlds by "carrying over" facts from the actual course of events which are not mentioned in the antecedent. The choice of these facts depends on whether they are *causally affected* by the truth or falsehood of the antecedent. This leads Mårtensson (1999) to postulate the following criterion of similarity between the world of evaluation and an alternative w':

Definition 43 (Causal trail criterion (Mårtensson, 1999))
If a conditional having A as antecedent is evaluated and A and C hold at w', then (intuitive) dissimilarity with respect to C should not count when assessing overall closeness iff the fact (at w') that C is in the causal trail of the fact (at w') that A.

Mårtensson leaves the notion of "causal trail" unanalyzed:

> ... I shall not attempt to define [the notion of causal trail] explicitly. Talk of a later event being in (or out of) the causal trail of some other event comes natural to me, and I hope (and trust) it will appear such to the reader too. (fn. 20, p. 102)

One may feel somewhat unsatisfied with this apparently cavalier treatment of a concept that is so central to the theory. After all, does not the use of counterfactuals "come natural" as well? Then why write a book about those? Such a reaction would be unfair, however. It is notoriously difficult to reduce causation to some other relation without circular reference to counterfactuals, a problem which has plagued theories since Goodman (1947). In Section 5.3 I will briefly discuss the notion, but the result will still be less than satisfactory.

A slightly refined version of the causal trail criterion is used in Mårtensson's treatment of the truth of counterfactuals. The refinements address two points: "Causal ramps" and "early departure."

5.2.4.2 Causal ramps

Lewis' (1979) "short transition period" before A holds at the alternative world, which Mårtensson follows Bennett in calling the "causal ramp," must still be provided for. In a world at which A is the case, this must have *come about* in some reasonable way. This gives rise to counterfactuals like (5.14):

(5.14) If a burglar were standing outside my apartment door right now, he would have set off the alarm at the gate.

Lewis pointed out that if the requirement that the time of divergence immediately precede the time of A is taken too literally, then

> ... it would be hard for me to say why some of this dependence should not be interpreted—wrongly, of course—as backward causation over short intervals of time in cases that are not at all extraordinary. (p. 40)

To avoid this counterintuitive result, the interpretation must "undo" not only the causal effects of the falsehood of A at this world, but also the earlier events that are responsible for A's falsehood.

Notice that according to Definition 43 the default is for true facts, posterior or otherwise, to be carried over *unless* they are causally affected by the antecedent. In fact, assuming as I am a nondeterministic world in which different futures do not imply different pasts, the time in question may be chosen to be the beginning of all time, if there is such a thing. The qualifying condition on alternative courses of history is that they agree with the actual world on all facts not causally affected by the antecedent. The preference for past predominance then falls out under the assumption that what is prior to A cannot be caused by A. What is needed to respect the causal ramp is that not *all* non-effects of A be asserted in the alternative course of history, but only those which did not, in our world, preempt A.

5.2.4.3 Early departure and disjunctive antecedents

It may not be possible for A to hold in an alternative history unless it is caused by some earlier event. If there are different alternatives in which sufficient causes occur at different times, all of them should be considered equally good candidates. Thus past predominance must again be relaxed to provide for such cases. This is illustrated by a well-known example attributed to John Pollock by Nute (1980):

> Suppose I left my coat in a classroom yesterday and was surprised to find it right where I left it today. Consider the antecedent
>
> (5.15) If my coat had been missing when I returned ...
>
> What true conditionals might have this antecedent? (p. 104)

The point is that on a strict interpretation of past predominance, it would follow, for instance, that a counterfactual with the consequent in (5.16) is true solely due to the temporal order of the possible causes of the antecedent:

(5.16) ... then I would sue John, for he was the only person in the room this morning and the coat wouldn't have been removed last night.

One way of approaching examples of this kind is to stipulate that the antecedent in question is in fact a *disjunction* of the form in (5.17a) which in turn is equivalent to the *conjunction* in (5.17b).

(5.17) a. If my coat had disappeared yesterday in the afternoon *or* last night *or* this morning *or* ...

b. If my coat had disappeared yesterday in the afternoon I would sue John *and* if it had disappeared last night I would sue John *and* if it had disappeared this morning I would sue John *and* ...

This solution is stipulative, however. Although it is a popular and in some cases reasonable assumption that counterfactuals with disjunctive antecedents should be equivalent to conjunctions in which the consequent is distributed over the disjuncts, that equivalence does not always hold.

(5.18) a. $(A \vee B) \rightarrow C$

b. $(A \rightarrow C) \wedge (B \rightarrow C)$

That the equivalence of (5.18a,b) appears valid in many cases is illustrated by the following examples, taken from the discussion in Chapter 4:

(5.19) a. If Jones had left through door A or through door C, he would have been killed.

b. If Jones had left through door A he would have been killed, and if he had left through door C he would have been killed.

However, Nute (1980, 1984) notes that if the inference is enshrined in an axiom, then the logic containing it must, if closed under substitution of provable equivalents, also endorse strengthening of the antecedent, among other undesirable consequences. This is easy to see: Let $A = (A \wedge B) \vee (A \wedge \overline{B})$; then if (5.18a,b) are equivalent, so are $A \rightarrow C$ and $((A \wedge B) \rightarrow C) \wedge ((A \wedge \overline{B}) \rightarrow C)$, hence $A \rightarrow C$ entails $(A \wedge B) \rightarrow C$.

There are also many linguistic examples which show that the equivalence is not a logical one, among them the following from Nute (1984):

(5.20) a. If the United States devoted more than half of its budget to defense or to education, it would devote more than half of its budget to defense.

b. If the United States devoted more than half of its budget to education, it would devote more than half of its budget to defense.

These examples are admittedly somewhat peculiar in that the antecedents are mutually incompatible and the consequent is one of them. However, they are clearly not infelicitous or hard to interpret.

The problem of disjunctive antecedents has been addressed in different ways, for instance by stipulating that the substitution with the conjunction sometimes is and sometimes is not warranted, or that the sentences of the different types have different logical forms (Fine, 1975; Loewer, 1976; Nute, 1980; Humberstone, 1978; Hilpinen, 1982, and others). While those accounts assign the sentences the intuitively correct interpretation, they lack any criteria for choosing between the alternatives for a given sentence. Others (Nute, 1984; Warmbrōd, 1981) prefer a pragmatic explanation according to which the set of accessible worlds depends on the surrounding context. Mårtensson (1999) gives the sentences different logical forms and attributes the applicability of the conversion to the number of different occasions on which the antecedent had a chance of being true.

It is of course the very point of employing non-monotonicity in the treatment of conditionals, be it via similarity between worlds or via probabilities, that only the most "prominent" among the ways in which the antecedents could have been true should be considered. Example (5.20) illustrates just that. I believe that the interpretation of examples like Pollock's involves a *backtracking* step which is not mentioned in the conditional itself, but is nevertheless indispensable. In the next section I will mention some more backtracking examples. For now, it is sufficient to make explicit the steps involved in the interpretation of the "coat" example:

(5.21) a. If my coat had been missing when I returned, John would have taken it.
b. If John had taken the coat, I would have sued him.
c. If my coat had been missing when I returned, I would have sued John.

(5.21) instantiates the probabilistically valid inference pattern of the form (5.22). However, whether the conclusion is in fact probable depends on the probabilities of *both* of (5.21a,b).

(5.22)
$$\frac{\begin{array}{c}A \to B\\ AB \to C\end{array}}{A \to C}$$

In Adams' probabilistic inference system, the probability of (5.21c) cannot be higher than that of (5.21a). It may be *possible* but not *likely* that John would have been the thief. Therefore whether (5.22c) follows from (5.22a) in conjunction with high probability of (5.22b) depends on the likelihood that it would have been John who took the coat.

Assuming a known set of possible ways in which the coat might have disappeared, for instance by being taken by either John or Pamela, the probability of (5.21a) is that of (5.23).

(5.23) If either John or Pamela had taken the coat, John would have taken it.

(5.23) is a case of *abduction*, that is, reasoning to the best explanation. If alternative explanations are conceivable, all of them need to be considered. This is the reason why past predominance is "overridden" in the case of the missing coat.

This concludes the informal discussion of the role of temporal and causal information in the interpretation of counterfactuals. For a discussion of how the causal trail criterion is related to various other notions such as *cotenability* (Goodman, 1947), *nomic pyramids* and *undercutting* (Pollock, 1976, 1981) and *despiteness* (Slote, 1978), see Mårtensson (1999).

5.3 Causality

The need for an account of causal relations has by now become evident both in the discussion of embedded conditionals in Chapter 4 and that of counterfactuals above. The upshot of Chapter 4 was that both the Import-Export Principle (Definition 31, page 95) and the Constancy condition (Definition 33, page 100) lead to counterintuitive predictions about the values of conditionals at the worlds where their antecedents are false. In the case of right-nested conditionals, the former makes better predictions than the latter. On the other hand, the Import-Export Principle is not available as a remedy for the case of conjunctions of conditionals.

In both cases, the predictions can be corrected by the same mechanism of adding restrictions to the antecedent. The same general strategy is suggested by the facts about counterfactuals discussed earlier in this chapter. In this section I am going to discuss the relevant background and propose an implementation in the interpretation function.

This section will not do justice to the ontological, metaphysical and epistemological issues surrounding the notion of causality.[6] Like objective chance, I take causal relations as given. The purpose of this section is not to analyze them, but to explain their formal representation. The questions that need to be addressed are: What are causes, what are effects, and what is the most economical way to represent them in the model?

5.3.1 What causal relations relate

Causal relations are usually stated as relations between types or tokens of *events*. That notion is somewhat ill-defined[7] and not sufficiently fine-grained, however. For instance, the event of my striking the match is identical to the event of my striking the match in the hallway, but it is only the former, not the latter, that caused the match to light. Only some "aspects" of the event are involved in the causal connection.

Hausman (1998) settles on *tropes*[8] to represent those aspects: "A trope is a located value of a variable or an instantiation of a property at a place and time." (p. 26) Tropes are, as it were, the "carriers" of the relation, the metaphysical basis for talk of both causal relations between *events* and explanatory relations between *facts*. I will exploit this connection between events and facts in encoding the relation in a model. Meanwhile I will continue to refer to the relation as one between events.

[6]For an excellent overview and arguments for the view I adopt, cf. Hausman (1998), especially Chapters 2 and 9.

[7]The "events" in statistics, which correspond to propositions in the current context, are unrelated to this usage, which has more in common with "events" as used in event semantics.

[8]For background, see Williams (1953); Armstrong (1997); Lewis (1983, 1986a, 1994). I will not discuss the differences and similarities between tropes and other entities such as states of affairs, truth-makers and situations.

5.3.2 Particular and general causation

Discussions of causality must distinguish relations between particular events (event *tokens*) from generalizations of those relations which hold among event *types*. The former is described with (5.24a), the latter with (5.24b).

(5.24) a. Her lack of exercise caused her heart attack.

b. Lack of exercise causes heart attack.

Superficially, pairs of descriptions such as (5.24a,b) appear to make the same claim. However, although the two are sometimes treated interchangeably (Suppes, 1970, is an example), that does not seem intuitively justified since the relations referred to have different properties. Causal relations between event tokens define a *strict partial order*, i.e., one that is

1. transitive (causal links form causal chains),
2. asymmetric (effects do not cause their causes), and
3. irreflexive (events do not cause themselves).

In contrast, the same does not appear to hold for event types, at least in common *descriptions* of relations between them. Davis (1988) discusses counter-evidence to all three of the above properties, involving examples like the following:

(5.25) a. An inability to run or walk causes one's leg muscles to become flaccid, and losing one's leg causes an inability to run or walk. Hence, losing one's leg causes one's leg muscles to become flaccid.

b. Jack and Jill continue to infect each other with colds: Their colds cause each other.

c. Violence begets violence.

(5.25a–c) arise from a "binding error" of sorts (a conventional one which does not render the descriptions infelicitous): None of the alleged relations is instantiated in a single tuple of event tokens. No case of inability to run or walk that is caused by leg loss causes the muscles to become flaccid. No single one of Jack's colds both causes and is caused by a single one of Jill's. No single act of violence begets itself.

It is true that causal relations between event tokens must somehow be related to generalizations over event types, as single-case chance must in general be related to relative frequencies. For the present purposes, however, only relations between events tokens are relevant, and only they will be represented in the model.

5.3.3 Deterministic causation of probabilities

The *'causes'*-relation explored in probabilistic theories is often paraphrased as *'increases the likelihood of (under certain circumstances)'*. Typically, the definition of the relation is some version or refinement of Definition 44, cited here after Davis (1988):

Definition 44 (Probabilistic causation)
C 'causes' E if and only if $Pr(E|C) > Pr(E)$ and there is no factor F such that $Pr(E|CF) = Pr(E|F)$.

In words, the probability that E occurs, given that C occurs, is greater than the unconditional probability that E occurs, and the increase in the probability of E in the presence of C is not independently brought about by some other factor.

Definition 44 is plagued by a number of problems. The probabilistic generalization is compatible with singular instances or small samples in which it is contradicted. Furthermore, it does not offer a general strategy for eliminating "spurious causes." These problems cast doubt mostly on the utility of the definition in the discovery and empirical verification of causal relations, a task which I am not concerned with here. I will only mention a *conceptual* problem arising in applying the definition to causal relations between singular event tokens.

E may fail to occur even though C occurred and *'caused'* E by raising its probability. This gives rise to paradoxical ways of speaking: The fact that my striking the match *'caused'* it to light is compatible with the fact that I struck it and it did not light. Furthermore, even if the match did light, the fact that this was a chance event which might have failed to occur despite the striking seems incompatible with the claim that it was caused (in the usual sense) by the striking.

These problems are avoided if causality is defined instead as a deterministic relation between events and probabilities. To adapt Hausman's example, "rather than regarding the match's lighting (if it does) as probabilistically caused, one should regard the probability that it lights as deterministically caused" (p. 201). Hausman lines up an array of further good arguments for this view. I adopt it here without further discussion.

5.3.4 Implementation

Events (and tropes) are located within worlds. They are not represented in a model as defined in Chapter 3, in which worlds have no structure other than a temporal succession of "snapshots." The model only represents the *facts* of worlds—true propositions which may be "about" particular times but are not located in time themselves.

Causal relations do not hold between facts. It is a fact that my striking the match caused it to light, but the fact that I struck it does not cause the fact that it lit. It *explains* the latter, but that is not the relation I am seeking to represent. Nevertheless, reference to facts is sufficient to state the causal relation without additional complexity in the model.

5.3.4.1 Causes

While a cause itself is part of a world, its occurrence is a fact about it, which can be stated to distinguish that world from others. The worlds in which it occurs form a proposition, which is identified by an indicator function. The language may contain a sentence that is true if and only if the cause is present, in which case the denotation of the sentence coincides with the random variable identifying the cause. There is no *a priori* reason to restrict the possible causes to those which can be referred to in this way, however.

5.3.4.2 Effects

An effect is not an event, but the probability of an event. The event itself, like its cause, is represented by the indicator function of the proposition that it occurs. The expectation of this indicator function is the chance that the event has of occurring. It can be read off the model as developed in Chapter 3: Each world at each time t is a member of the set $h(t)$ for some history h. The chance of the event is $Pr(\text{'the event occurs'}|h(t))$.

Since the probability and its development over time are explicitly coded, the statement of the causal relation need not include the numerical value. It is instead defined solely in terms of the random variables involved: A statement of the form 'X causally affects Y', where X and Y are random variables, asserts that the *value* of X causally affects the *expectation* of Y; alternatively, the expectation of Y causally depends on the value of X.

5.3.4.3 Definitions

The last section concluded with an informal statement of the 'causally affects'-relation. As I said above, only particular causation is represented. The relation is transitive, asymmetric and irreflexive.[9] Definition 45 fixes the notion of a *causal structure* as a strict partial order on an arbitrary (and possibly empty) set.

[9] Transitivity is commonly assumed for particular causation (Davis, 1988, p. 146), but I should emphasize that it only holds for the relation itself, *not* necessarily for the probabilistic "impulse" transmitted along a causal chain; that impulse may be canceled out by the values of intermediary variables; cf. also Hausman (1998, p. 204); Pearl (2000, p. 237).

Definition 45 (Causal structure)
A causal structure *is a pair* $\langle \Phi, \prec \rangle$*, where* Φ *is a set and* $\prec \subseteq \Phi \times \Phi$ *is a relation which is*

a. *transitive: For all* $\varphi, \varphi', \varphi'' \in \Phi$*, if* $\varphi \prec \varphi'$ *and* $\varphi' \prec \varphi''$*, then* $\varphi \prec \varphi''$*;*

b. *asymmetric: For all* $\varphi, \varphi' \in \Phi$*, if* $\varphi \prec \varphi'$ *then* $\varphi' \not\prec \varphi$*; and*

c. *irreflexive: For all* $\varphi \in \Phi$*,* $\varphi \not\prec \varphi$*.*

The idea is that $\varphi \prec \varphi'$ holds whenever φ *'causally affects'* φ' in the above sense. In the interpretation of conditionals, in fact, it will be more useful to identify those variables which are *'causally unaffected'* by a given variable.

Definition 46 (Causal independence)
Given a causal structure $\langle \Phi, \prec \rangle$*, for all* $\varphi, \varphi' \in \Phi$*:* φ' *is* causally independent *of* φ *if and only if* $\varphi \not\prec \varphi'$*.*

An ordinary chance model can now be augmented with a causal structure on the indicator functions of a set of propositions.

Definition 47 (Causal chance model)
A causal chance model *is a structure* $\langle W, T, <, \approx, Pr, \Phi, \prec \rangle$*, where* $\langle W, T, <, \approx, Pr \rangle$ *is a chance model (cf. Definition 17, page 72),* Φ *is a set of functions in* $\{0,1\}^W$*, and* $\langle \Phi, \prec \rangle$ *is a causal structure.*

In Definition 47, a subset of all random variables is singled out and collected in Φ. Technically there is no need to give them such prominence; $\prec$ could instead be defined as a relation on the set of *all* random variables. It is largely for notational convenience that I treat them separately.

There is no limitation on the number of variables that can stand in causal relations, and I believe that there is no such limitation in the real world either, if only for the reason that causal chains can be dense. I suspect, however, although I have no empirical evidence for this, that if the evaluation in the model is to serve as a realistic reflection of the *cognitive* task of interpreting conditionals, the set of variables that are considered relevant in any given instance should be rather sparse.

5.4 Applications

Returning now to the task of assigning values to conditionals at the worlds where their antecedents are false, the addition of causal information to the model allows us to impose restrictions on the set of "visible" alternative worlds over which the expectation of the consequent is taken. The idea is simple and characterized informally in (5.26).

(5.26) Only those worlds are accessible in which the variables that are causally independent of the antecedent have the same value that they have at the world of evaluation.

Formally, in the framework for right-nested conditionals of Chapter 4, the interpretation can be made sensitive to causal relations by modifying the Simplification map (Definition 36, page 104) as follows:

Some examples will help to illustrate.

5.4.1 Right-nested conditionals

Consider first the case of first-order and right-nested predictive conditionals.

(5.27) a. If the match is wet, then if you strike it, it will light.

$W^{t'} \to (S^{t'} \to \mathbf{F}L)$

b. If you strike the match, it will light.

$S^{t'} \to \mathbf{F}L$

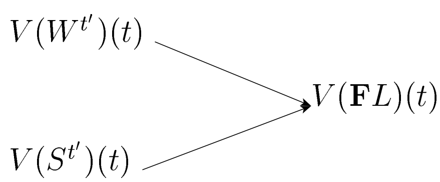
The formal representation assumes that both $W^{t'}$ and $S^{t'}$ refer to the same time. In fact, *'the match is wet'* may refer to the present instead, making (5.27a) an epistemic conditional. That difference is not crucial here; the question of which alternatives to make accessible from those worlds where the match is not wet pertains to both.

It is reasonable to assume that the condition of the match is not causally affected by whether you strike it. On the other hand, both the condition of the match and whether you strike it causally affect its lighting. The causal structure relevant for the interpretation of the sentence is given in (5.28a,b); the order is displayed graphically in (5.28c).[10]

(5.28) a. $\Phi = \{V(W^{t'})(t), V(S^{t'})(t), V(\mathbf{F}L)(t)\}$

b. $V(W^{t'})(t) \prec V(\mathbf{F}L)(t)$ and $V(S^{t'})(t) \prec V(\mathbf{F}L)(t)$

c.



What is important for the interpretation of the conditional is that the wetness and the striking are causally independent. According to the informal characterization in (5.26), the wetness of the match must remain "intact" in the selection of alternative worlds at which the match is struck.

Consider a world at which the match is wet and you do not strike it. The value of the consequent of (5.27a) is the expectation of it lighting, taken over the "visible" alternatives at which you strike it. Since the fact that it is wet is causally independent of the striking, only "wet" worlds are visible. The value of the consequent is small.

Likewise, from worlds at which the match is dry at the relevant time and you do not strike it, only those worlds are visible at which you strike it and it is dry. The value of the consequent is accordingly high. In summary, the values of (5.27b) are assigned according to the rule in (5.29).

[10] Arrows indicate the direction of the *'causally affects'*-relation. The pictorial device of *directed acyclic graphs* is preferred in some formal accounts of causality, notably Pearl (2000). Pearl invariably represents the background factors (corresponding here to the history up to the time in question) as an additional variable.

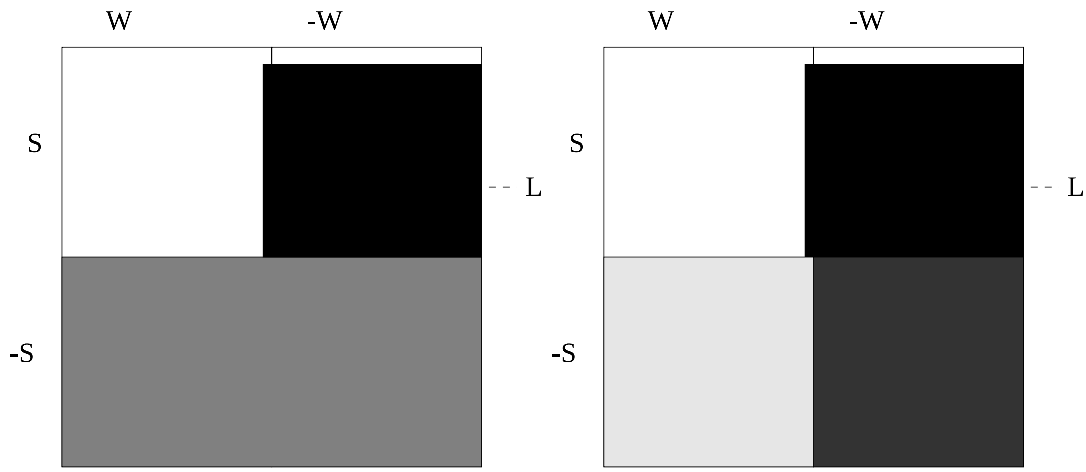


Figure 5.1: The distribution of the values of $V(S \to L)$ without (left) and with (right) causal restrictions

(5.29) $$V(S^{t'} \to \mathbf{F}L)(t)(w) = \begin{cases} V(\mathbf{F}L)(t)(w) \text{ if } V(S)(t')(w) = 1 \\ E[V(\mathbf{F}L)(t) | V(S)(t') = 1, V(W)(t') = x] \\ \qquad \text{if } V(S)(t')(w) = 0 \text{ and } V(W)(t')(w) = x \end{cases}$$

Figure 5.1 illustrates the distribution of the values of (5.28b) without (left) and with (right) consideration of the condition of the match. The values are indicated with colors: black for 1, white for 0, and grey for intermediate values. The values are listed in (5.30a–d).

(5.30) a. 1 at worlds where the match is struck and lights;

b. 0 at worlds where it is struck and does not light;

c. $P(\mathbf{F}L|S^{t'}W^{t'})$ at worlds where it is wet and not struck (lower left);

d. $P(\mathbf{F}L|S^{t'}\overline{W}^{t'})$ at worlds where it is dry and struck (lower right).

The overall expectation of these values is still the probability that it lights, given that it is struck, as the proof of Fact 8 spells out in tedious detail.

Fact 8
$P(S^{t'} \to \mathbf{F}L|h(t)) = P(\mathbf{F}L|S^{t'}, h(t))$.[11]

While the uneven distribution of the values of (5.27b) over the worlds where $S^{t'}$ is false does affect the probability of that sentence, it does make a difference for the probability of (5.27a). According to the general rule, the values of (5.27b) are assigned at those worlds where $W^{t'}$ is true, and the expectation over all worlds at which $W^{t'}$ is true is assigned at those worlds where $W^{t'}$ is false. That expectation depends on whether the wetness of the match was taken into account in the calculation of the values of (5.27b), as illustrated in Figure 5.2.

More explicitly, the values assigned are as given in (5.31).

(5.31) $V(W^{t'} \to (S^{t'} \to \mathbf{F}L))(t)(w)$

$$= \begin{cases} V(S^{t'} \to \mathbf{F}L)(t)(w) \text{ if } V(W)(t')(w) = 1 \\ E[V(S^{t'} \to \mathbf{F}L)(t)|V(W)(t') = 1] \text{if } V(W)(t')(w) = 0 \end{cases}$$

The expectation of the values is derived in some detail the proof of Fact 9.

---

[11] *Proof.* For readability, values at worlds where the match is struck are underlined as far as they are kept separate. Also for readability, I drop temporal indices and the restriction to $h(t)$, and I abbreviate '$V(\varphi) = 1$' as 'φ.'

$$\begin{aligned} P(S \to L) =& E[V(S \to L)] = \underline{E[V(S \to L)|S']Pr(S')} + E[V(S \to L)|\overline{S}']Pr(\overline{S}') \\ =& \underline{Pr(L'|S')Pr(S')} \\ & + E[V(S \to L)|SW']Pr(\overline{S}W') + E[V(S \to L)|S\overline{W}']Pr(\overline{S}\,\overline{W}') \\ =& \underline{Pr(SL')} \\ & + Pr(L'|SW')Pr(\overline{S}W') + Pr(L'|S\overline{W})Pr(\overline{S}\,\overline{W}') \\ =& \underline{Pr(L'|SW')Pr(SW') + Pr(L'|S\overline{W}')Pr(S\overline{W}')} \\ & + Pr(L'|SW')Pr(\overline{S}W') + Pr(L'|S\overline{W}')Pr(\overline{S}\,\overline{W}') \\ =& Pr(L'|SW')\big[Pr(SW') + Pr(\overline{S}W')\big] \\ & + Pr(L'|SW')\big[Pr(\overline{S}W') + Pr(\overline{S}\,\overline{W}')\big] \\ =& Pr(L'|SW')Pr(W') + Pr(L'|S\overline{W}')Pr(\overline{W}') \\ =& Pr(L'|S') = P(L|S) \end{aligned}$$

□

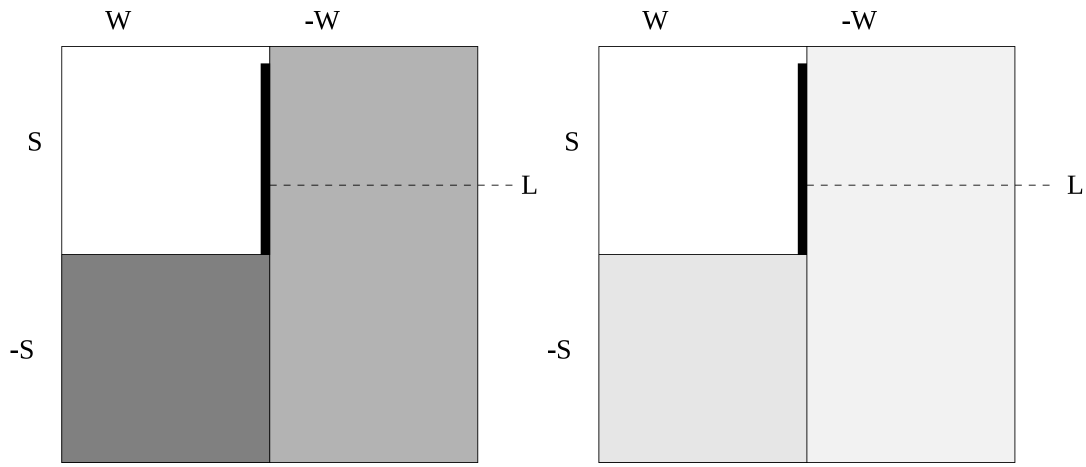


Figure 5.2: The distribution of the values of $V(W \to (S \to L))$ without (left) and with (right) causal restrictions

Fact 9

$$P(W^{t'} \to (S^{t'} \to \mathbf{F}L)|h(t))$$
$$= P(S^{t'}\mathbf{F}L|W^{t'}, h(t)) + \underline{P(\mathbf{F}L|S^{t'}W^{t'}, h(t))}P(\overline{S}^{t'}|W^{t'}, h(t)).$$ [12]

This is not the value of Jeffrey (1991), which was criticized by Edgington (1995) (cf. Section 4.4.4.1, page 110) and is repeated here as (5.32).

(5.32) $$P(W^{t'} \to (S^{t'} \to \mathbf{F}L)|h(t))$$
$$= P(S^{t'}\mathbf{F}L|W^{t'}, h(t)) + \underline{P(\mathbf{F}L|S^{t'}, h(t))}P(\overline{S}^{t'}|W^{t'}, h(t))$$

[12] *Proof.* Again, simplifying the notation somewhat:

$$\begin{aligned}
&P(W \to (S \to L)) \\
&\quad = E[V(W \to (S \to L))|W']Pr(W') + E[V(W \to (S \to L))|\overline{W}']Pr(\overline{W}') \\
&\quad = \big(Pr(SL'|W') + Pr(L'|SW')Pr(\overline{S}'|W')\big)Pr(W') \\
&\quad\quad + E[V(W \to (S \to L))|W']Pr(\overline{W}') \\
&\quad = Pr(SL'|W') + Pr(L'|SW')Pr(\overline{S}'|W')\big[Pr(W') + Pr(\overline{W}')\big] \\
&\quad = P(SL|W) + P(L|SW)P(\overline{S}|W)
\end{aligned}$$

□

The difference lies in the underlined terms, representing the values at worlds at which the match is wet and not struck.

Thus for the right-nested conditional, the effect taking the causal independence of the wetness and the striking into account is similar to that of the Import-Export Principle, but the claim is that this is due to the role of causal relationships, not to the form of (5.31a): Even the values of (5.31b) at worlds where the match is not struck vary depending on the condition of the match.

5.4.2 Conjunctions of conditionals

The preceding example showed that by taking causal relations into account one can mimic the effect of the Import-Export Principle. Therefore it did not illustrate the invalidity of the Principle. A more interesting example in this regard is the case of conjunctions. There is no need to discuss it at the same level of detail as the preceding one, since the two are similar in relevant respects.

(5.33) If Jones left through door a he was killed, and if he left through door c he was killed.

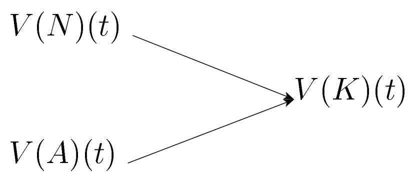
I will represent the event that the werewolf is stalking neighborhood M as $\overline{N}$. I will drop the temporal indices since (5.33) is an epistemic conditional in which all constituents refer to the time of evaluation.

The causal relations involved are similar to the preceding example: The werewolf's and Jones' whereabouts are causally independent, and both causally affect his chance of being eaten. The relevant structure is depicted in (5.34).

(5.34) a. $\Phi = \{V(N)(t), V(A)(t), V(K)(t)\}$

b. $V(N)(t) \prec V(K)(t), V(A)(t) \prec V(K)(t)$

c.



These causal relations imply that in assigning values to the conditionals at the non-antecedent worlds, the whereabouts of the werewolf at those worlds need to be left "intact." At worlds at which Jones left through door a and was killed (thus the werewolf is stalking n) the value of $C \to K$ is the expectation that he was killed, given that he left through door c and the werewolf is stalking n.

$$(5.35) \qquad V(A \to K)(t)(w) = \begin{cases} V(K)(t)(w) \text{ if } V(A)(t)(w) = 1 \\ E[V(K)(t)|V(A) = 1, V(N) = x] \\ \qquad \text{if } V(A)(t)(w) = 0 \text{ and } V(N)(t)(w) = x \end{cases}$$

The expectation of the values assigned according to (5.35) is given in Fact 10. I already argued in Section 4.4.4.2 (page 113) that this is the desired value, not that of (5.36).

Fact 10
$P((A \to K)(C \to K)) = \dfrac{P(K|AN)P(CK) + P(K|CN)P(AK)}{P(A \vee C)}$.[13]

$$(5.36) \qquad P((A \to K)(C \to K)) = \frac{P(K|A)P(CK) + P(K|C)P(AK)}{P(A \vee C)}$$

5.4.3 Counterfactuals: Prior propensity

The paraphrases I used to justify the values at non-antecedent worlds repeatedly made reference to intuitions about counterfactuals. With the formal implementation of causal relations, this reference can now be made more precise. That there is a

[13]Proof omitted.

close relation between the values of predictive conditionals and counterfactuals is illustrated in (5.37a,b).

(5.37) a. If Oswald does not shoot Kennedy, somebody else will.
b. If Oswald had not shot Kennedy, somebody else would have.

We can assume (at least for the sake of argument) that Oswald did indeed shoot Kennedy and that he acted alone. The intuition shared by most speakers is that both of (5.37a,b) are false and, moreover, that whatever made (5.37a) false before November 22, 1963, also makes (5.37b) false now.

The simplest way to account for this intuitive connection is to suppose that (5.37b) is merely the past tense of (5.37a). Proposals to this effect have been made by a number of authors (e.g., Adams, 1975; Slote, 1978; Skyrms, 1981; Tedeschi, 1981; Ellis, 1984; Pendlebury, 1989; Dudman, 1994; Edgington, 1995; Dahl, 1997; Dancygier, 1998).[14] The basic approach is known under different labels reflecting the overall convictions of the respective exponents. Skyrms (1981) calls it the *prior propensity* account, while Adams (1975) prefers to talk about *epistemic past*. Barker (1998) refers to it as *tense probabilism*. I believe that the approach is on the right track. It has come under criticism, however, and I need to clarify in what sense it is right.

Barker (1998) noted that one cannot consistently maintain that the probabilities of the two sentences are equal and that the world is non-deterministic. This is illustrated by examples like those I used in the section on the relevance of posterior facts in interpreting counterfactuals (Section 5.2.3.2) Recall the example of the coin toss (page 125), repeated here as (5.38).

(5.38) a. Before the toss: If I bet on tails, I will lose.
b. After the toss: If I had bet on tails, I would have lost.

After the toss, in light of the fact that the coin came up heads, (5.38b) is true. However, before the toss, the probability that the coin would come up heads was .5

[14]The problems I am about to discuss were noticed, but not addressed in any detail, by some of these authors.

(assuming that the coin is fair), hence (5.38a) was not true. To suppose otherwise would require the quite unwarranted assumption that the outcome was somehow already determined before the toss. Such an assumption would be at odds both with naïve intuitions about the scenario and with the metaphysical commitment to objective chance and non-determinism underlying the prior propensity approach. So, Barker argues, while the fact that the coin came up heads is taken into account in interpreting the counterfactual, it is irrelevant to the predictive conditional, therefore tense probabilism is not viable as a unified account of conditionals.

Notice, however, that all that is demonstrated by this argument is that the two sentences are not (in general) *equiprobable*; it does not follow that they are not *equivalent*. In the random-variable approach, two numbers must not be confused: the *values* assigned to the conditional at individual worlds, and the *probability* of the conditional (i.e., the expectation of its values). According to the definitions given here, considering the causal relations at work in (5.38a,b), the values of the indicative at non-antecedent worlds are the same as those of the counterfactual; their probabilities, however, differ.

The outcome of the coin toss is causally independent of my betting heads or tails, and both the outcome of the toss and my betting determine whether I win or lose. Thus the causal relations are again as in the earlier examples; I will not reiterate the technical details in this section.

Consider a world at which I do not bet on tails and the coin comes up heads. In evaluating, at a time t before the toss, the predictive conditional in (5.38a), those alternatives at which I bet on tails are examined. The fact that the coin comes up heads is not causally affected by my betting, hence only those worlds are "visible" at which I bet on tails and the coin comes up heads, and so I lose.

Thus (5.38a) is true at such a world. Its probability at t, however, is not high: There are alternatives in $h(t)$ in which I do not bet on tails and the coin comes up tails. From such worlds, only those are antecedent-worlds visible at which I bet on tails and the coin still comes up tails, so I win and (5.38a) is false. These alternatives are responsible for the low expectation of the values of the predictive conditional.

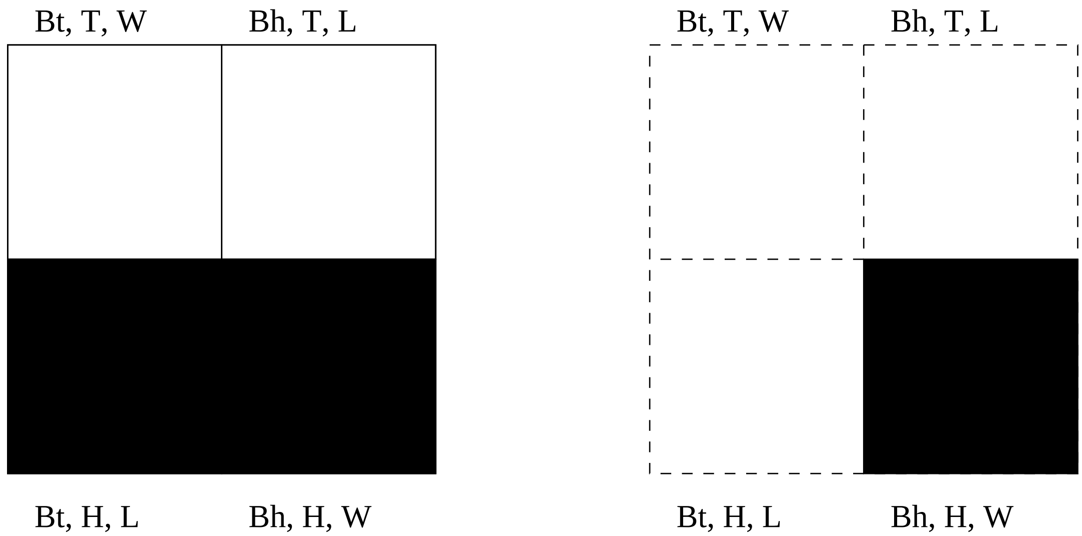


Figure 5.3: Distribution of the values of (5.38a) before the toss (left) and those of of (5.38b) after the speaker bet heads and the coin came up heads (right). 'Bt/Bh' = '*bet tails/heads*'; 'H/T = '*comes up heads/tails*'; 'W/L' = '*win/lose*'.

In short, once again the values of the predictive conditional are not distributed uniformly over those worlds at which its antecedent is false. Figure 5.3 (left) illustrates their distribution. The antecedent '*If I bet tails* ... ' is true in the two quadrants on the left. The values of '*If I bet tails I will lose*' at worlds where I bet heads (two quadrants on the right) differ according as the coin comes up tails (top) or heads (bottom).

As history progresses, some historical alternatives are discarded, and at a later time, in $h(t')$, only those remain at which the coin came up heads and the value of the predictive conditional was accordingly high. This is illustrated on the right in Figure 5.3, where only the worlds in the lower quadrant right quadrant remain. Consequently, the counterfactual is felt true even though the earlier prediction is not.

Formally, this suggests the logical form in (5.39b) for the sentence in (5.39a).[15]

(5.39) a. If I had bet on tails, I would have lost.

[15]A representation of this sort was suggested by Thomason and Gupta (1981).

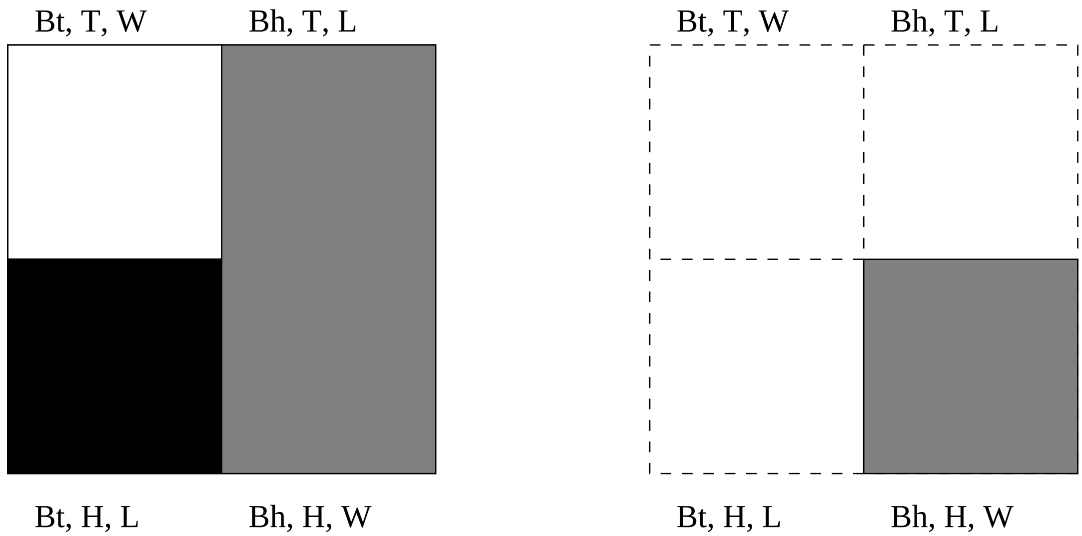


Figure 5.4: Distribution of the values of (5.38a) before the toss (left) and those of of (5.38b) after the speaker bet heads and the coin came up heads (right) assuming that the bet causally affects the toss.

b. $P(B^{t'} \to \mathbf{F}W)$

The sentence under the scope of the past operator is a typical predictive conditional. The values assigned to (5.39b) at the later time t' are the values assigned to its predictive counterpart at t at those worlds where the coin comes up heads. As I argued above, those values are higher than the expectation of $B^{t'} \to \mathbf{F}W$ at t.

Finally, suppose the coin toss is *not* causally independent of the betting: If I bet on heads, you toss the coin with your left hand; if I bet on tails, you toss it with your right hand. I bet on heads, you tossed it with your left hand, and it came up heads. Now, unlike in the earlier scenario, (5.40b) is not very likely: If I had bet on tails, you would have tossed the coin with the your other hand and it might have come up tails.

(5.40) a. If I bet on tails, I will lose.

b. If I had bet on tails, I would have lost.

The difference between these two scenarios is properly accounted for by the value assignment. Since the coin toss, and hence its outcome, depends on the bet, it, too, must be "undone" in making antecedent-worlds visible from those in which I bet on heads. Now the value assigned at those worlds is taken over *all* the worlds in which I bet tails, and it is accordingly low. This is illustrated in Figure 5.4. Notice that the expectation of the predictive conditional (5.40a) before the toss is unaffected by the change in the scenario, which, I believe, is intuitively correct.

5.4.4 Import-Export revisited

The last examples I want to discuss here shed some light on the intuitive plausibility of the Import-Export Principle and its logical non-validity. I used (5.41a) to argue against the Principle. Suppose t' precedes t''. Intuitively, the sentence is not necessarily true or even highly probable. In particular, in a setting in which the match may light without being struck (because it is tossed into the fire instead), it may be highly likely that it will light while it is not likely that it will burn if struck.

(5.41) a. If the match lights, then it will light if you strike it.
$L^{t''} \rightarrow (S^{t'} \rightarrow \mathbf{F}L)$

b. If the coin comes up heads, then it will come up heads if you bet on tails.
$H^{t''} \rightarrow (B^{t'} \rightarrow \mathbf{F}W)$

For (5.41b), in contrast, there does not seem to be similar scenario. The more likely it is that the coin will come up heads, the more likely it is that it will come up heads (and you lose) if you bet on tails.

The paraphrases in (5.41a,b) bring out both the temporal relations and the semantic difference a bit clearer.

(5.42) a. If it is true (now) that the match lights at t'', then it is also true (now) that if you strike it at t', it will light.

b. If it is true (now) that the coin comes up heads at t'', then it is also true (now) that if you bet on tails at at t', it will come up heads.

There is no obvious way to attribute the difference to either the form of the sentences or the temporal relations. Causal relations make the right predictions in both cases.

Whether the match lights causally depends on whether you strike it. Whether the coin comes up heads is independent of which way you bet. The difference is reflected in the assignment of values at which the respective embedding antecedent is true while the embedded antecedent is false.

From a world at which the match lights without you striking it, all the worlds at which you strike it are "visible," since the lighting of the match must be "undone" in determining the set of alternatives. In contrast, from a world at which the coin comes up heads and you bet on heads, only some of the worlds at which you bet on tails are visible: those at which the coin, in addition, comes up heads. That latter fact is not undone.

5.5 Summary

This chapter clarified the relationship between the values of predictive and epistemic conditionals at worlds where their antecedents are false and the values of counterfactuals. Causal relations, which have been known to affect the interpretation of counterfactual conditionals, are as important for non-counterfactuals. Taking them into account leads to intuitively correct predictions about the probabilities of compounded and embedded conditionals. Furthermore, the random variable approach, made sensitive to causal relations in this way, explains the intuition behind the prior propensity account of counterfactuals: Predictive conditionals "before the fact" and counterfactuals "after the fact" are equivalent, but not equiprobable.

Chapter 6

Conclusion

The semantics of conditionals has long been an open issue, and it is likely to remain open for some time to come. This dissertation explored the application of the formal technique of random variables and intermediate values to epistemic, predictive and counterfactual conditionals. I discussed its motivation and its intuitive rationale, and I suggested ways of overcoming its shortcomings. I believe that its capability of linking truth and probability in a straightforward and elegant way and its potential to provide a unified account of at least these three kinds of conditionals make it an exciting and promising object of research.

I left many questions unanswered, however, some by choice and others because I do not have the answers at this time. For instance, I did not discuss the values of conditionals with conditional antecedents. That the formalism provides such values is known and I present the required model-theoretic apparatus in the appendix. I also believe that those values are intuitively correct. But it is not easy to have intuitions about such sentences, and a discussion of the predicted values would have been rather removed from linguistic analysis.

Perhaps less excusable is my failure to give a full account of certain uses of the conditionals that are within the scope of this work, such as “backtracking” counterfactuals and conditionals with disjunctive antecedents. They are the subject of ongoing work which I will present on a later occasion.

Other conditionals which I did not touch upon in this work may be susceptible to a probabilistic analysis as well. I expect this to be the case for generic conditionals and not for conditional questions and commands, but I do not yet have a definite opinion at this point.

More generally, the introduction of probabilistic methods into semantic analysis raises a number of related questions. What, for instance, is the relationship between uncertainty about word denotations and uncertainty about sentence denotations? What would a compositional probabilistic semantics be like, one which would incorporate the linguistic sources of uncertainty that I chose to ignore? Can the predictions of a probabilistic semantic theory be tested experimentally rather than by introspection?

This dissertation scratched the surface of a big and multifaceted problem. How deep a dent I have made is not for me to judge.

Appendix A

Free compounds and embeddings

A.1 Introduction

The chapters so far have only dealt with "first-order" and right-nested conditionals and conjunctions thereof. The last step is to extend the assignment of values to the full language containing conditionals with conditional antecedents. I saved this extension for the last chapter because it requires a substantial complication of the model. I am not aware of a simple assignment rule, not even one involving a syntactic transformation, which would implement the idea as straightforwardly. That is not to say that there is no such rule; I leave the search for it for future work. Aside from the treatment of conditional antecedents it affords, the model-theoretic extension is instructive in its own right, because it will provide a general role which applies to right-nested conditionals and conjunctions thereof as well and makes the assignment defined earlier fall out naturally.

I do not claim credit for inventing the approach; it is adopted directly from van Fraassen (1976) and Stalnaker and Jeffrey (1994).

A.2 Informal characterization

The problem with conditional antecedents is that their antecedents do not denote propositions, hence nothing to condition on. This problem did not arise for right-nested conditionals whose antecedents were guaranteed to be truth-functional.

In Section 2.6, page 57, I briefly sketched van Fraassen's motivation for the assignment of intermediate values: Consider a conditional like (A.1) and a world w.

(A.1) $\varphi \to C$

If φ is a truth-functional sentence, it has a truth value at w. If it is true, the value of the conditional is that of C. Where φ is false, a φ-world w' is chosen *at random*. If C is true (false) at w', the conditional is true (false) at w. The *expectation* over all possible such trials is just the probability of the consequent conditional on the antecedent, and it is the intermediate value assigned at w.

The above simplifying paraphrase presupposes that φ has truth-values at all worlds. This is not the case when it is itself a conditional, as in (A.2).

(A.2) $(A \to B) \to C$

The antecedent $A \to B$ has truth values at the A-worlds, but it is also true "to some extent" at the $\overline{A}$-worlds. This blurs the distinction between those worlds at which the antecedent is false and for which an alternative must be chosen, and those where it is true and which are therefore eligible alternatives.

I know of no general repair for this problem in the simple models I dealt with so far. van Fraassen (1976) proposed an ingenious way out of this dilemma.

A.3 Implementation

We need the distribution of values over a set $h(t)$, given a probability distribution $Pr(\cdot|h(t))$. For notational convenience, I will omit the reference to $h(t)$ and simply refer to the set as W and to the distribution as Pr, taking the simple model $\langle W, Pr \rangle$ as the point of departure. From $\langle W, Pr \rangle$ we construct a structure $\langle W^*, Pr^* \rangle$ in which all sentences of the language, including conditionals, denote propositions.

Definition 48 (Stalnaker Bernoulli Model (van Fraassen, 1976))
A Stalnaker Bernoulli Model *based on a model* $\langle W, Pr \rangle$ *is a structure* $\langle W^*, Pr^* \rangle$*, where*

a. W^* *is the set of denumerable sequences of worlds in* W*, represented as functions* $w^* : \mathbb{N} \mapsto W$*;*

b. *for all* $X_1 \times \ldots \times X_n \times W^* \subseteq W^*$*, where* $X_1, \ldots, X_n \subseteq W$*,*[1]

$$Pr^*(X_1 \times \ldots \times X_n \times W^*) = Pr(X_1) \cdot \ldots \cdot Pr(X_n).$$

The idea is that each "world" w^* in W^* represents an infinite sequence of trials, picking worlds from the "urn" W with replacement.[2] The "propositions" in the product space W^* are Cartesian products of propositions in W. For instance, $X \times \overline{X} \times W^*$ is the event that an X-world is drawn first and a $\overline{X}$-world next, the set of all sequences w^* such that $w^*(1) \in X$ and $w^*(2) \notin X$. Also, $W \times X \times W^*$ is the event that an X-world is chosen on the second trial, and so on.

Of particular importance for present purposes is the event that an X-world is reached at *some* point in the sequence of trials. Proposition 49 states that whenever X has non-zero probability, however small, an X-world *will* be reached.

Proposition 49
If $Pr(X) > 0$*, then the event that an* X*-world is reach in a sequence* $w^* \in W^*$ *has probability one.*[3]

[1] Some authors (van Fraassen, 1976; Stalnaker and Jeffrey, 1994) avoid the assumption made here that Pr is defined for *all* subsets of W, using instead a privileged field of propositions as the *generating sets* for $\langle W^*, Pr^* \rangle$. The construction defined here, where all subsets of W are generating sets, is a special case of that.

[2] Hence van Fraassen's term "Stalnaker Bernoulli Model."

Proposition 49 is important because it will allow us to neglect those sequences which consist entirely of $\overline{X}$-worlds. What that means and why it should be helpful will become clear momentarily. First I define the assignment of values to truth-functional sentences for the new model.

Each proposition X in the original set W corresponds to the set $X \times W^*$, and the definition of the product measure Pr^* ensures that $Pr^*(X \times W^*) = Pr(X)$. Notice also that what remains after "lopping off" the first world from a sequence w^* is again an infinite sequence of worlds, hence a member of W^*. I will write w^{*++} for the sequence starting with the second world in w^*; thus for all n, $w^{*++}(n) = w^*(n+1)$.

The assignment function is now extended to give sentences truth values at world sequences $w^* \in W^*$. The values of truth-functional sentences are defined solely by reference to the first member $w^*(1)$. Conditionals are treated as follows: If the antecedent is true at w^*, the value of the conditional is that of the consequent at w^*. Otherwise, the value is that of the conditional at w^{*++}. By recursion, this value is obtained from the final sub-sequence starting with $w^*(n)$ for the smallest n such that the antecedent is true at $w^*(n)$.

Definition 50 (Assignment function)
Given an assignment function $V^1 : \mathcal{L}^1_\mathcal{A} \mapsto \{0,1\}^W$ for a model $\langle W, Pr\rangle$, an assignment function $V^* : \mathcal{L}_\mathcal{A} \mapsto \{0,1\}^{W^*}$ is defined for the corresponding Stalnaker Bernoulli Model $\langle W^*, Pr*\rangle$ as follows: For all $w^* \in W^*$ and $\varphi \in \mathcal{L}_\mathcal{A}$,

$$\textit{If } \varphi \in \mathcal{L}^1_\mathcal{A} : V^*(\varphi)(w^*) = V^1(\varphi)(w^*(1))$$

$$V^*(\varphi \to \psi)(w^*) = \begin{cases} V^*(\psi)(w^*) \textit{ if } V^*(\varphi)(w^*) = 1 \\ V^*(\varphi \to \psi)(w^{*++}) \textit{ otherwise} \end{cases}$$

---

[3] *Proof.*
$$\begin{aligned} & Pr^*(\{w^* \in W^* | w^*(n) \in X \text{ for some } n \in \mathbb{N}\}) \\ &= 1 - Pr^*(\{w^* \in W^* | w^*(n) \notin X \text{ for all } n \in \mathbb{N}\}) \\ &= 1 - \lim_{n\to\infty} Pr(\overline{X})^n \\ &= \begin{cases} 1 \text{ if } Pr(X) > 0; \\ 0 \text{ otherwise} \end{cases} \end{aligned}$$
□

The following properties of Definition 50 are worth noting: Firstly, only truth values are assigned. No expectations, hence no intermediate values, arise here. As a consequence, conditionals are propositions where their values are defined. Secondly, those values are undefined at sequences which do not contain any world at which the antecedent is true. Thirdly, no special provisions are made for more complex conditionals.

It is with respect to the second point that Proposition 49 is of great use: For any conditional $A \to B$, the probability that no A-world is ever reached, hence the value of $V^*(A \to C)$ undefined in a sequence, is zero. Hence the expectation of $V^*(A \to B)$ equals the conditional expectation, given that $V^*(A \to B)$ is defined. The complement set of those sequences can be ignored. This extends to sentences $\varphi \to \psi$ containing an arbitrary finite number of conditionals.

As for the third point, the fact that conditionals have truth values wherever they have values at all, and that the event that they do not have values has zero probability, allows for a treatment of complex conditionals on a par with simple ones. Each constituent can be expected to have a truth value (unless an antecedent has zero probability.)

Let us consider the set of all sequence at which a conditional $A \to C$ is true. This is the set $\{w^* \in W^* | V^*(A \to C)(w^*) = 1\}$; for the moment, I will refer to it simply as $A \to C'$. The probability $Pr^*(A \to C')$ is given by the assignment from Definition 50. Recall that Pr^* is defined as a product measure in terms of Pr, which is defined for the propositions in the original model $\langle W, Pr\rangle$. How, then, is $Pr^*(A \to C')$ related to Pr? Since conditionals do not in general denote propositions in $\langle W, Pr\rangle$, $Pr^*(A \to C')$ is not equivalent to $Pr(X)$ for any $X \subseteq W$. Intuitively, this is because $A \to C'$ "cuts across" bundles of sequences with the same first member at which $V(A)$ is false.

However, $Pr^*(A \to C')$ is the *product* of probabilities defined in $\langle W, Pr\rangle$. To utilize this fact, the "Fraction Lemma" of van Fraassen (1976) is helpful.

Lemma 51 (Fraction Lemma (van Fraassen, 1976, p. 294))
For all X such that $Pr(X) \neq 0$, $\sum_{i=0}^{\infty} Pr(\overline{X})^i = 1/Pr(X)$.[4]

The probability $Pr(A \to C')$ is now easily obtained:

Theorem 52

For all A, C, $Pr(A \to C') = Pr(C'|A')$.[5]

The probabilities of more complex conditionals are defined in $\langle W^*, Pr^* \rangle$, but cannot be compared against $\langle W, Pr \rangle$, which provides only simple conditional probabilities. Instead, I proceed to define the values assigned by the random variable at each of the worlds in the original model $\langle W, Pr \rangle$, now in terms of the Stalnaker Bernoulli model $\langle W^*, Pr^* \rangle$.

The value $V(\varphi)(w)$ for a sentence φ at a world w will be the expectation of $V^*(\varphi)$ over a certain set of world sequences. If φ has a determinate truth value at w, the expectation should be that value. Otherwise, the expectation should be the correct intermediate value.

Notice that due to the product construction of W^*, if the first member is "lopped off" all sequences beginning with the same world w, the set of remaining sub-sequences is a faithful copy of the full space W^* (A.3):

(A.3) $$\{w^{*++}|w^*(1) = w\} = W^*$$

---

[4]*Proof.* $Pr(X)\sum_{i=0}^{\infty} Pr(\overline{X})^i$
$= \sum_{i=0}^{\infty} Pr(\overline{X})^i Pr(X)$
$= \sum_{i=0}^{\infty} Pr(\{w^* \in W^* | w^*(n) \notin X \text{ for all } n < i \text{ and } w^*(i) \in X\})$
$= Pr(\{w^* \in W^* | w^*(n) \in X \text{ for some } n\})$
$= 1$ (by Proposition 49 and since $Pr(X) \neq 0$.)
□

[5]*Proof.* For each $w^* \in W^*$, $V^*(A \to C)(w^*)$ is determined by the smallest n such that $V(A)(w^*(n)) = 1$. The set $\{w^* \in W^* | V^*(A \to C)(w^*) = 1\}$ is a union of disjoint sets of sequences; for each such set there is some n such that for all of its members w^*, $V(A)(w^*(i)) = 0$ for all $i < n$ and $V(AC)(w^*(n)) = 1$. Thus
$Pr^*(A \to C') = Pr(AC') + Pr(\overline{A}')Pr(AC') + Pr(\overline{A}')^2 Pr(AC') + \ldots$
$= Pr(AC') \cdot \sum_{i=0}^{\infty} Pr(\overline{A}')^i$
$= Pr(AC')/Pr(A') = Pr(C'|A')$ (by Lemma 51.)
□

This suggests a straightforward definition for the value assignment to worlds in W:

Definition 53
For all $w \in W$ and $\varphi \in \mathcal{L}_{\mathcal{A}}$:

$$V(\varphi)(w) = Pr^*(\{w^{*++} | w^*(1) = w \text{ and } V^*(\varphi)(w^*) = 1\})$$

If φ does not contain conditionals, then $V^*(\varphi)(w \times W^*) = V(\varphi)(w)$. If $\varphi = A \rightarrow C$ and $V(A)(w) = 1$, then $V(\varphi)(w) = V(C)(w)$, as before. However, if $V(A)(w) = 0$, then $V^*(\varphi)$ is true at some of the sequences starting with w and false at others. By Theorem 52, the measure of the set of sequences at which it is true is just the conditional probability.

Bibliography

Adams, E. 1965. The logic of conditionals. *Inquiry*, 8:166–197.

Adams, E. 1970. Subjunctive and indicative conditionals. *Foundations of Language*, 6:89–94.

Adams, E. 1975. *The Logic of Conditionals*. Reidel.

Adams, E. 1998. *A Primer of Probability Logic*. CSLI Publications.

Anderson, A. R. 1951. A note on subjunctive and counterfactual conditionals. *Analysis*, 12:35–38.

Appiah, A. 1984. Jackson on the material conditional. *Australasian Journal of Philosophy*, 62(1):77–81.

The Aristotelian Society. 1991. *The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the University of Durham*. Supplementary Volume 65.

Armstrong, D. M. 1997. *A World of States of Affairs*. Cambridge University Press.

Athanasiadou, A. and R. Dirven, editors. 1997. *On Conditionals Again*, volume 143 of *Amsterdam Studies in the Theory and History of Linguistic Science*. John Benjamins, Amsterdam / Philadelphia.

Barker, S. 1998. Predetermination and tense probabilism. *Analysis*, 58:290–296.

Barwise, J. 1986. Conditionals and conditional information. In Traugott et al. (1986), pages 21–54.

Bennett, J. 1984. Counterfactuals and temporal direction. *The Philosophical Review*, 93(1):57–91.

Bennett, J. 1988. Farewell to the phlogiston theory of conditionals. *Mind*, 97(388): 509–527.

Bennett, J. 1995. Classifying conditionals: The traditional way is right. *Mind*, 104 (414):332–354.

Braithwaite, R., editor. 1931. *The Foundations of Mathematics and other Logical Essays.* Routledge and Kegan Paul, London.

Burgess, J. P. 1979. Logic and time. *The Journal of Symbolic Logic*, 44(4):566–582.

Carlstrom, I. and C. Hill. 1978. Review of Adams 1975. *Philosophy of Science*, 45: 155–158.

Carnap, R. 1945. The two concepts of probability. *Philosophy and Phenomenological Research*, 5:513–532.

Chisholm, R. 1946. The contrary-to-fact conditional. *Mind*, 55:289–307.

Chuaqui, R. 1991. *Truth, Possibility and Probability.* Number 166 in North-Holland Mathematics Studies. North-Holland, Amsterdam / New York / Oxford / Tokyo.

Comrie, B. 1985. *Tense.* Cambridge University Press, Cambridge / New York.

Comrie, B. 1986. Conditionals: A typology. In Traugott et al. (1986), pages 77–99.

Dahl, O. 1997. The relation between past time reference and counterfactuality: A new look. In Athanasiadou and Dirven (1997), pages 97–112.

Dancygier, B. 1998. *Conditionals and Prediction.* Cambridge University Press.

Davis, W. A. 1988. Probabilistic theories of causation. In Fetzer (1988), pages 133–160.

van Deemter, K. and S. Peters, editors. 1996. *Semantic Ambiguity and Underspecification.* Number 55 in CSLI Lecture Notes. CSLI, Stanford, CA.

Downing, P. B. 1959. Subjunctive conditionals, time order, and causation. *Proceedings of the Aristotelian Society*, 59:125–140.

Dudman, V. 1984a. Conditional interpretations of *If*-sentences. *Australian Journal of Linguistics*, 4:143–204.

Dudman, V. 1984b. Parsing 'If'-sentences. *Analysis*, 44(4):145–153.

Dudman, V. 1988. Indicative and subjunctive. *Analysis*, 48(3):113–122.

Dudman, V. 1989. Vive la Révolution! *Mind*, 98(392):591–603.

Dudman, V. 1991. Interpretations of 'if'-sentences. In Jackson (1991), chapter 10, pages 202–232.

Dudman, V. 1994. On conditionals. *Journal of Philosophy*, 91:113–128.

Edgington, D. 1986. Do conditionals have truth conditions? *Critica*, 18(52):3–30. Reprinted in Jackson (1991), pages 176–201.

Edgington, D. 1991. The mystery of the missing matter of fact. In *The Aristotelian Society* Ari (1991), pages 185–209. Supplementary Volume 65.

Edgington, D. 1995. On conditionals. *Mind*, 104(414):235–329.

Edgington, D. 1997. Commentary. In *Conditionals*, pages 97–137. Clarendon Press, Oxford, UK.

Eells, E. and B. Skyrms, editors. 1994. *Probabilities and Conditionals: Belief Revision and Rational Decision*. Cambridge University Press, Cambridge.

Ellis, B. 1978. A unified theory of conditionals. *Journal of Philosophical Logic*, 7: 107–124.

Ellis, B. 1984. Two theories of indicative conditionals. *Australasian Journal of Philosophy*, 62(1):50–66.

Fetzer, J. H., editor. 1988. *Probability and Causality*, volume 192 of *Studies in Epistemology, Logic, Methodology, and Philosophy of Science*. D. Reidel, Dordrecht / Boston / Lancaster / Tokyo.

Fine, K. 1975. Review of Lewis (1973). *Mind*, 84:451–458.

Fleischman, S. 1989. Temporal distance: A basic linguistic metaphor. *Studies in Language*, 13(1):1–50.

Fowler, H. 1950. *A Dictionary of Modern English Usage*. Oxford University Press, New York.

van Fraasen, B. C. 1981. A temporal framework for conditionals and chance. In Harper et al. (1981), pages 323–340.

van Fraassen, B. C. 1976. Probabilities of conditionals. In Harper and Hooker (1976), pages 261–308.

Funk, W.-P. 1985. On a semantic typology of conditional sentences. *Folia Linguistica*, 19(3/4):365–414.

Gabbay, D. and F. Guenthner, editors. 1984. *Extensions of Classical Logic*, volume 2 of *Handbook of Philosophical Logic*. D. Reidel.

Gärdenfors, P. 1988. *Knowledge in Flux: Modeling the Dynamics of Epistemic States*. MIT Press.

Gibbard, A. 1981. Two recent theories of conditionals. In Harper et al. (1981), pages 211–247.

Goodman, N. 1947. The problem of counterfactual conditionals. *The Journal of Philosophy*, 44:113–128. Reprinted in Jackson (1991), pages 9–27.

Grandy, R. and R. Warner, editors. 1986. *Philosophical Grounds of Rationality*. Clarendon Press, Oxford.

Grice, H. 1989. *Studies in the Way of Words*. Harvard University Press.

Grünwald, P. 1998. *The Minimum Description Length Principle and Reasoning under Uncertainty*. PhD thesis, ILLC, Amsterdam.

Hájek, A. 1994. Triviality on the cheap? In Eells and Skyrms (1994), pages 113–140.

Hájek, A. and N. Hall. 1994. The hypothesis of the conditional construal of conditional probability. In Eells and Skyrms (1994), pages 75–110.

Hall, N. 1994a. Back in the CCCP. In Eells and Skyrms (1994), pages 141–160.

Hall, N. 1994b. Correcting the guide to objective chance. *Mind*, 103:505–518.

Harper, W. and C. Hooker, editors. 1976. *Foundations of Probability Theory, Statistical Inference, and Statistical Theories of Science*, volume 1 of *The University of Western Ontario Series in Philosophy of Science*. D. Reidel.

Harper, W., R. Stalnaker, and G. Pearce, editors. 1981. *Ifs: Conditionals, Belief, Decision, Chance, and Time*. Reidel.

Hausman, D. M. 1998. *Causal Asymmetries*. Cambridge University Press, Cambridge, UK.

Hilpinen, R. 1982. Disjunctive permissions and conditionals with disjunctive antecedets. In Niiniluoto, I. and E. Saarinen, editors, *Proceedings of the Second Soviet-Finish Logic Conference*. Acta Philosophica Finnica, Moscow.

Humberstone, I. L. 1978. Two merits of the circumstantial operator language for conditional logic. *Australasian Journal of Philosophy*, 56:21–24.

Iatridou, S. 2000. The grammatical ingredients of counterfactuality. *Linguistic Inquiry*, 31(2):231–270.

Jackson, F. 1979. On assertion and indicative conditionals. *Philosophical Review*, 88: 565–589. Reprinted in Jackson (1991), pages 111–135.

Jackson, F. 1984. Two theories of indicative conditionals: Reply to Brian Ellis. *Australasian Journal of Philosophy*, 62(1):67–76.

Jackson, F. 1987. *Conditionals*. Basil Blackwell, Oxford.

Jackson, F., editor. 1991. *Conditionals*. Oxford University Press.

James, D. 1982. Past tense and the hypothetical: A cross-linguistic study. *Studies in Language*, 6(3):375–403.

Jeffrey, R. C. 1964. If. *Journal of Philosophy*, 61:702–703.

Jeffrey, R. C., editor. 1980. *Studies in Inductive Logic and Probability*, volume 2. University of California Press.

Jeffrey, R. C. 1991. Matter-of-fact conditionals. In *The Symposia Read at the Joint Session of the Aristotelian Society and the Mind Association at the University of*

Durham, pages 161–183. The Aristotelian Society. Supplementary Volume 65.

Joos, M. 1964. *The English Verb*. University of Wisconsin Press, Madison, WI.

Karttunen, L. and S. Peters. 1979. Conventional implicature. In Oh, C.-K. and D. Dinneen, editors, *Presupposition*, volume 11 of *Syntax and Semantics*, pages 1–56. Academic Press, New York.

Keefe, R. and P. Smith, editors. 1996. *Vagueness: A Reader*. MIT Press, Cambridge, MA / London, U.K.

Kratzer, A. 1979. Conditional necessity and possibility. In Egli, U., R. Bäuerle, and A. von Stechow, editors, *Semantics from Different Points of View*, pages 117–147. Springer.

Kratzer, A. 1981. Partition and revision: The semantics of counterfactuals. *Journal of Philosophical Logic*, 10:201–216.

Kratzer, A. 1986. Conditionals. In Farley, A. M., P. T. Farley, and K.-E. McCullough, editors, *Proceedings of CLS 22, Part 2: Papers from the Parasession on Pragmatics and Grammatical Theory*, pages 1–15.

Lance, M. 1991. Probabilistic dependence among conditionals. *Philosophical Review*, 100:269–276.

Lewis, D. 1973. *Counterfactuals*. Harvard University Press.

Lewis, D. 1976. Probabilities of conditionals and conditional probabilities. *Philosophical Review*, 85:297–315. Reprinted with a Postscript in Jackson (1991), pages 76–101. Page numbers refer to Jackson (1991).

Lewis, D. 1979. Counterfactual dependence and time's arrow. *Noûs*, 13:455–476. Reprinted with Postscripts in Lewis (1986b), pages 32–66, and Jackson (1991); pages 46–75. Page numbers refer to Lewis (1986b).

Lewis, D. 1980. A subjectivist's guide to objective chance. In Jeffrey (1980), pages 263–293. Reprinted in Harper et al. (1981), pages 267–297; with a Postscript in Lewis (1986b), pages 83–132. Page numbers refer to Lewis (1986b).

Lewis, D. 1981a. Causal decision theory. *Australasian Journal of Philosophy*, 59: 5–30. Reprinted with a Postscript in Lewis (1986b), pages 305–339. Page numbers refer to Lewis (1986b).

Lewis, D. 1981b. Ordering semantics and premise semantics for counterfactuals. *Journal of Philosophical Logic*, 10(2):217–234. Reprinted in Lewis (1998), pages 77–96. Page numbers refer to Lewis (1998).

Lewis, D. 1983. New work for a theory of universals. *Australasian Journal of Philosophy*, 61:343–377. Reprinted in Lewis (1999), pages 8–55. Page numbers refer to Lewis (1999).

Lewis, D. 1986a. Against structural universals. *Australasian Journal of Philosophy*, 64:25–46. Reprinted in Lewis (1999), pages 78–107. Page numbers refer to Lewis (1999).

Lewis, D. 1986b. *Philosophical Papers*, volume 2. Oxford University Press.

Lewis, D. 1986c. Probabilities of conditionals and conditional probabilities II. *Philosophical Review*, 95:581–589. Reprinted in Jackson (1991), pages 102–110; Lewis (1998), pages 57–65. Page numbers refer to Lewis (1998).

Lewis, D. 1994. Humean Supervenience debugged. *Mind*, 103:473–490. Reprinted in Lewis (1999), pages 224–247. Page numbers refer to Lewis (1999).

Lewis, D. 1998. *Papers in Philosophical Logic*. Cambridge University Press.

Lewis, D. 1999. *Papers in Metaphysics and Epistemology*. Cambridge University Press, Cambridge.

Loewer, B. 1976. Counterfactuals with disjunctive antecedents. *Journal of Philosophical Logic*, 73:531–536.

Mårtensson, J. 1999. *Subjunctive Conditionals and Time: A Defense of the Classical Approach*. Number 10 in Acta Philosophica Gothoburgensia. Department of Philosophy, University of Göteborg.

McGee, V. 1989. Conditional probabilities and compounds of conditionals. *The Philosophical Review*, 98(4):485–541.

McGee, V. 1994. Learning the impossible. In Eells and Skyrms (1994), pages 179–199.

Mellor, D., editor. 1978. *Foundations: Essays in Philosophy, Logic, Mathematics and Economics*. Routledge and Kegan Paul, London.

Mellor, D., editor. 1990. *Philosophical Papers: F.P. Ramsey*. Cambridge University Press, Cambridge, U.K.

Mellor, D. 1993. How to believe a conditional. *The Journal of Philosophy*, 90:233–248.

Milne, P. 1997. Quick triviality proofs for probabilities of conditionals. *Analysis*, 57: 75–80.

Nelson, E. 1987. *Radically Elementary Probability Theory*. Number 117 in Annals of Mathematical Studies. Princeton University Press, Princeton, NJ.

Nute, D. 1980. *Topics in Conditional Logic*. D. Reidel.

Nute, D. 1984. Conditional logic. In Gabbay and Guenthner (1984), pages 387–439.

Palmer, F. 1983. Future time reference in the conditional protasis: A comment on Comrie. *Australian Journal of Linguistics*, 3:241–243.

Pearl, J. 2000. *Causality: Models, Reasoning, and Inference.* Cambridge University Press.

Pendlebury, M. 1989. The projection strategy and the truth conditions of conditional statements. *Mind*, 98(390):179–205.

Pinkal, M. 1985. *Logik und Lexikon: Die Semantik des Unbestimmten.* de Gruyter, Berlin. Translated as Pinkal (1995).

Pinkal, M. 1995. *Logic and Lexicon: The Semantics of the Indefinite.* Kluwer, Dordrecht / Boston. Translated by Geoffrey Simmons.

Poesio, M. 1996. Semantic ambiguity and perceived ambiguity. In van Deemter and Peters (1996), pages 159–201.

Pollock, J. L. 1976. *Subjunctive Reasoning.* Reidel.

Pollock, J. L. 1981. A refined theory of counterfactuals. *Journal of Philosophical Logic*, 10:239–266.

Portner, P. 1992. *Situation Theory and the Semantics of Propositional Expressions.* PhD thesis, University of Massachusetts, Amherst, MA.

Prior, A. 1967. *Past, Present and Future.* Oxford University Press.

Quirk, R., S. Greenbaum, G. Leech, and J. Svartvik. 1985. *A Comprehensive Grammar of the English Language.* Longman, London / New York.

Ramsey, F. 1929. General propositions and causality. Printed in Braithwaite (1931), pages 237–255; Mellor (1978), pages 133–151; Mellor (1990), pages 145–163. Page numbers refer to Mellor (1990).

Reichenbach, H. 1956. *The Direction of Time.* University of Berkeley Press, Berkeley and Los Angeles, CA.

Roberts, C. 1989. Modal subordination and pronominal anaphora in discourse. *Linguistics and Philosophy*, 12:683–721.

Russell, B. 1923. Vagueness. *Australasian Journal of Philosophy and Psychology*, 1: 84–92. Reprinted in Keefe and Smith (1996), pages 61–68. Page numbers refer to this reprint.

Skyrms, B. 1980. *Causal Necessity: A Pragmatic Investigation of the Necessity of Laws.* Yale University Press, New Haven and London.

Skyrms, B. 1981. The prior propensity account. In Harper et al. (1981), pages 259–265.

Skyrms, B. 1984. *Pragmatics and Empiricism.* Yale University Press, New Haven / London.

Skyrms, B. 1988. Conditional chance. In Fetzer (1988), pages 161–178.

Slote, M. A. 1978. Time in counterfactuals. *The Philosophical Review*, 87(1):3–27.

Sosa, E., editor. 1975. *Causation and Conditionals*. Oxford University Press.

Stalnaker, R. 1968. A theory of conditionals. In *Studies in Logical Theory, American Philosophical Quarterly, Monograph: 2*, pages 98–112. Blackwell. Reprinted in Sosa (1975), pages 165–179; Harper et al. (1981), pages 41–55; Jackson (1991), pages 28–45. Page numbers refer to Harper et al. (1981).

Stalnaker, R. 1970. Probablity and conditionals. *Philosophy of Science*, 37:64–80. Reprinted in Harper et al. (1981), pages 107–152. Page numbers refer to Harper et al. (1981).

Stalnaker, R. 1975. Indicative conditionals. *Philosophia*, 5:269–286. Reprinted in Jackson (1991), pages 136–154; and Stalnaker (1999), pages 63–77. Page numbers refer to Stalnaker (1999).

Stalnaker, R. 1976. Letter to van Fraassen. In Harper and Hooker (1976), pages 302–306.

Stalnaker, R. 1981. A defense of conditional excluded middle. In Harper et al. (1981), pages 87–104.

Stalnaker, R. 1991. Letter to Richard Jeffrey. Included in Jeffrey (1991), pages 178–183.

Stalnaker, R. 1999. *Context and Content*. Oxford Cognitive Science Series. Oxford University Press, Oxford.

Stalnaker, R. and R. Jeffrey. 1994. Conditionals as random variables. In Eells and Skyrms (1994), pages 31–46.

von Stechow, A. and D. Wunderlich, editors. 1991. *Semantik: Ein internationales Handbuch der zeitgenössischen Forschung. [= Semantics]*. Walter de Gruyter.

Strawson, P. 1986. ‘If’ and ‘$\supset$’. In Grandy and Warner (1986), pages 229–242.

Suppes, P. 1970. *A Probabilistic Theory of Causality*. North Holland.

Tedeschi, P. 1981. Some evidence for a branching-futures semantic model. In Tedeschi, P. and A. Zaenen, editors, *Tense and Aspect*, volume 14 of *Syntax and Semantics*, pages 239–270. Academic Press, New York.

Thau, M. 1994. Undermining and admissibility. *Mind*, 103:491–503.

Thomason, R. H. 1970. Indeterminist time and truth value gaps. *Theoria*, 36:264–281.

Thomason, R. H. 1984. Combinations of tense and modality. In Gabbay and Guenthner (1984), pages 135–165.

Thomason, R. H. and A. Gupta. 1981. A theory of conditionals in the context of branching time. In Harper et al. (1981), pages 299–322.

Tichy, P. 1976. A counterexample to the Lewis-Stalnaker analysis of counterfactuals. *Philosophical Studies*, 29:271–273.

Traugott, E. C., A. ter Meulen, J. Snitzer Reilly, and C. A. Ferguson, editors. 1986. *On Conditionals*. Cambridge University Press.

Veltman, F. 1985. *Logics for Conditionals*. PhD thesis, University of Amsterdam.

Warmbrōd, K. 1981. Counterfactuals and substitution of equivalent antecedents. *Journal of Philosophical Logic*, 10(2):267–289.

Williams, D. 1953. The elements of being. *Metaphysical Review*, 7:3–18 and 171–192.